Solving Professional Problems Together: The Impact of Collaboration on Pre-Service Teachers' Scientific Reasoning

Dissertation zum Erwerb des Doctor of Philosophy (Ph.D.)

am Munich Center of the Learning Sciences

der Ludwig-Maximilians-Universität München

Vorgelegt von

Andras Csanadi

München, 27.04.2017

Erstgutachter: Prof. Dr. Ingo Kollar

Zweitgutachter: Prof. Dr. Frank Fischer

Datum der Mündlichen Prüfung: 14.06.2017

Acknowledgements

I would like to express my gratitude towards those without whom this dissertation would not have been made. Namely, I would like to thank to my supervisors (Ingo Kollar and Frank Fischer), David Shaffer and the Epistemic Analytics Group, my Reason-Fellows and my Family.

The research presented in this work was supported by the Elite Network of Bavaria [Project number: K-GS-2012-209]. I would like to extend my sincere gratitude to the opportunities (conference attendance, incubator stay, international knowledge exchange) made possible by the ENB.

Content

Abstract

Future professionals should be prepared for scientific reasoning, i.e., to construct and apply scientific knowledge, in order to analyze and solve problems in their professional practice. Yet, future practitioners' scientific reasoning skills often seem to be deficient when solving practical problems. This dissertation explores to what extent collaboration may foster the engagement of future practitioners in scientific reasoning: i.e., in epistemic processes (e.g., hypothesizing, evaluating evidence) and in referring to scientific content knowledge (e.g., scientific theories and evidence). Therefore, two studies were conducted to compare collaborative and individual problem solving of pre-service teachers regarding their scientific reasoning. Study 1 investigates the effect of group heterogeneity with respect to problem solving scripts on scientific reasoning. Study 2 explores to what extent Epistemic Network Analysis can serve as a methodological approach for measuring scientific reasoning. As part of Study 1, pre-service teachers solved an educational problem either as individuals (N=16) or as pairs (N=30 pairs). Collaboration showed a mixed effect on scientific reasoning processes: pairs engaged more in explaining and reasoning about the problem and drew more conclusions, while individuals engaged more in generating solutions. Additional analyses showed that the more heterogeneous pairs were regarding their members' problem solving scripts the more they engaged in hypothesizing and evaluating evidence and the less they engaged in generating solutions. Finally, pairs less often referred to scientific content than individuals did during problem solving. Study 2 further analyzed the data by applying Epistemic Network Analysis. This method has the advantage of analyzing patterns of connections between epistemic processes, i.e., epistemic networks of scientific reasoning. The central epistemic process for pairs was evidence evaluation, which they frequently used in combination with hypothesizing and communicating and scrutinizing. On the other hand, the most characteristic process in individuals' scientific reasoning was solution generation, which very often co-occurred with hypothesizing and evidence evaluation. The overall results indicate that if the aim is to develop a more reflective understanding of the problem, future practitioners should collaborate with each other, especially in heterogeneous settings. However, they should be supported (1) to share knowledge regarding scientific theories and evidence as well as (2) to reach a mutual understanding (e.g., by coordinating explanations) on the problem after a certain time so as to be able to have the capacity of generating solutions. Moreover, the different effects of collaboration on the process and content aspects of scientific reasoning imply that scientific reasoning might not be a unidimensional construct, and its process and content levels should be differentiated in future research. A further important methodological implication is that the process aspect of scientific reasoning can be analyzed as a network of interconnected skills and such analysis might bring more explanatory value than the mere reliance on frequencies of occurrences of isolated activities.

Keywords: Scientific Reasoning, Epistemic Processes, Collaborative Problem Solving, Group Heterogeneity, Epistemic Network Analysis

1. Problem statement

Not only for scientists it is essential to know how to construct and apply valid knowledge in a reliable manner in order to examine and solve problems in the context of their professional practice. For – partly or even predominantly – practice-oriented professions such as medicine (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996), business (Stark, Gruber, Mandl, & Hinkofer, 2001) or teaching it is also important that practitioners know how to utilize relevant knowledge, so as their professional decisions follow a systematic reasoning process and are made on reliable evidence. This means that when they solve problems in their practice, practitioners should know how to reason scientifically, i.e., to (1) engage in problem solving in a systematic way analogous to scientific reasoning (Fischer et al., 2014); and (2) use adequate scientific sources (e.g., theoretical constructs or research findings) from their domain (e.g., Spencer, Detrich, & Slocum, 2012). For example, doctors solve diagnostic problems on a daily basis (Charlin, Boshuizen, Custers, & Feltovich, 2007). In order to cure their patients they should (1) systematically collect information (e.g., from the patient and from medical records or by initiating further investigations) so as to test plausible hypotheses (e.g., about illnesses) as well as (2) apply their academic knowledge (e.g., about diseases) in order to be able to professionally explain and alleviate the patients' problem. To give another example, a teacher might notice that one or more of her pupils underperforms on a test. This teacher should be able to critically reflect on this phenomenon, in an evidence-based manner (Robinson, 1998; Spencer et al., 2012; Voss, Kunter, & Baumert, 2011). This means that first, the teacher should be capable of engaging in certain problem solving or reasoning processes (e.g., identifying the problem, evaluating her own assumptions). Second, she should refer to theoretical knowledge on learning and instruction (e.g., considering cognitive, emotional, social or institutional factors that may influence the students' performance) as well as she may need to relate that knowledge to empirical findings (e.g., research on epistemic motives or learning strategies). While solving problems as scientifically knowledgeable practitioners (i.e., reasoning scientifically; Fischer et al., 2014) is generally part of the curricula in higher education (similarly, in teacher education), the difficulties of utilizing scientific reasoning skills for solving complex practice-related problems is a frequently reported issue (Gräsel & Mandl, 1993; Gruber, Mandl, & Renkl, 2000). For example, medical students often show problems to transfer the knowledge they have learnt in the university context (i.e., at lectures) to professionally engage in diagnostic problem solving (Gräsel & Mandl, 1993). Similarly, teachers often seem to rely to a limited extent on scientific resources when handling professional problems (Hetmanek et al., 2015), and they often seem to show difficulties in using evidence to assess learning and instruction (Morris, 2006; Yeh & Santagata, 2015).

This dissertation argues that one possibility to engage future practitioners in scientific reasoning while solving professional problems is to ask them to solve such problems collaboratively during their academic training (Baeten & Simons, 2016; Noroozi, Teasley, Biemans, Weinberger, & Mulder, 2013; Rummel & Spada, 2005) rather than to deal with them alone. Collaboration can be seen an authentic context of scientific reasoning and argumentation (Osborne, 2010; Simon, Langley, & Bradshaw, 1981) and, therefore, collaborative reasoning (e.g., solving problem in pairs) might "naturally" trigger an engagement in scientific reasoning compared to individual reasoning that may be a less natural setting for reasoning scientifically. The main assumption behind engaging future professionals more often in collaborative problem solving practices during their academic training is that it could offer them the opportunity to learn (1) how to solve problems professionally (i.e., in a science-based manner) and (2) content knowledge in a more authentic way (e.g., Brown, Collins, & Duguid, 1989) which can further facilitate the transferability of these knowledge to their future professional problem solving practice (Hmelo-Silver, 2004; Stark, Gruber, Mandl, & Renkl, 1998). To engage future practitioners in collaboration so as it is beneficial for their scientific reasoning, it should also be considered, however, that there are factors reported to facilitate or hinder group productivity (Hill, 1982; Hirst & Echterhoff, 2012; Noroozi et al., 2013; Paulus, 2000). These factors might, similarly, have an impact on practitioners' engagement in scientific reasoning during collaboration. One can account for many of these by giving further instructional support (e.g., Mullins, Rummel, & Spada, 2011; e.g., Noroozi et al., 2013), while there are only some that one may consider without necessarily giving further instructions. Since the scope of this dissertation is to investigate the impact of collaboration on scientific reasoning without further instructional support, it seems reasonable to set a focus on factors that reasoners bring to the problem solving context and that might have an impact on whether collaboration may be a beneficial learning context for learning how to reason scientifically. Prior knowledge might be an ideal candidate for such purpose, because it is shown to affect reasoning processes (Shapiro, 2004; Stark, Puhl, & Krause, 2009). To give an example, if one is planning to optimize collaboration between future practitioners, she can create a group of reasoning partners who are, for example, more or less diverse in their prior knowledge. Group diversity (e.g., based on prior knowledge) is widely cited among those factors that may influence collaborative reasoning (Bowers, Pharmer, & Salas, 2000; Fischer & Mandl, 2005; Noroozi et al., 2005; Paulus, 2000; Rummel & Spada, 2005). Therefore, it is an interesting question to further investigate whether a certain type of prior knowledge can be identified to affect practitioners' engagement in collaborative scientific reasoning. For example, collaborative partners' diverse approaches on problem solving, i.e., the heterogeneity of their "problem solving scripts" (Fischer, Kollar, Stegmann, & Wecker, 2013; Schank, 1999) might influence how they engage in collaborative scientific reasoning. Although there is an empirical need to explore how such scripts may influence reasoning processes (Vogel, Wecker, Kollar, & Fischer, 2016), there seems to be a research gap in this area. Therefore, it is a further aim of this dissertation to explore a potential impact of the heterogeneity of groups regarding their problem solving scripts on their engagement in collaborative scientific reasoning.

Furthermore, to find valid answers regarding the earlier argument that collaboration might be a more authentic context and, therefore, more advantageous to scientific reasoning compared to individual reasoning; selecting the appropriate methodological approach is important. Such a method should offer interpretable and comparable representations of scientific reasoning processes of collaborating partners as well as of individuals. Authors often propose that scientific reasoning is a complex process (e.g., Osborne, 2010) that can be measured through different subskills or processes (Chinn & Malhotra, 2002; Fischer et al., 2013; Klahr & Dunbar, 1988). Yet, beyond theoretical proposals for such processes (e.g., Fischer et al., 2013) there is a lack of methodological framework on measuring scientific reasoning in a conclusive manner. As a result, studies often apply "coding and counting" methods (Okada & Simon, 1997) and measure scientific reasoning processes in an "isolated" manner without being able to draw valid conclusions about the overall process of scientific reasoning. Therefore, a further aim of this dissertation is to find a methodological framework that accounts for individual processes contributing to scientific reasoning and, at the same time, allows conclusive interpretations regarding scientific reasoning.

To sum up, the main aim of this dissertation is to investigate the question whether collaboration might be an effective way to deal with the problem of future professionals' suboptimal engagement in scientific reasoning. In this respect, the following work investigates (1) what role diverse (i.e., heterogeneous) group composition might play in collaborative scientific reasoning; and (2) what could be an optimal methodological approach to assess scientific reasoning while comparing groups with individuals. For the proper investigation of these questions, further specification on what constitutes scientific reasoning is necessary. Therefore, the next chapter (Chapter 2) is concerned with the conceptualization of scientific reasoning in its relation to practitioners' problem solving. The subsequent chapter (Chapter 3) addresses the main question of the study, i.e., how a collaborative context may affect scientific reasoning, and discusses theoretical and empirical sources to find existing answers to that question. The chapters afterwards consider the two sub-problems related to the main question of this work: (1) the group composition problem (Chapter 4) and (2) the measurement problem (Chapter 5-6). After setting the research questions (Chapter 7) for further inquiry, two empirical studies are introduced to find answers to the main question on the effect of collaboration (Chapter 8) as well as on the sub-questions on group-composition (Chapter 8) and on measurement (Chapter 9). Finally, the last chapters (Chapter 10-11) inspect the generalizability and the limitations of the empirical findings reported in the earlier chapters (Chapter 8-9).

2. Scientific reasoning

In order to answer the main question of this dissertation whether collaboration can be a beneficial context to engage future practitioners in scientific reasoning while they solve problems relevant for their future practice, a conceptual framework on scientific reasoning is necessary. Therefore, in the following, a theoretical framework is introduced in three steps. First, a short review will investigate the commonalities and the distinctive features between scientific reasoning and problem solving; so as to clarify if the conceptual differentiation between scientific reasoning and problem solving can be warranted (Chapter 2.1). The subchapter afterwards argues how scientific reasoning is relevant for solving professional problems in practice (Chapter 2.2). The subsequent subchapters (Chapter 2.3-2.5) introduce a two-dimensional framework on scientific reasoning. Finally, a review of empirical literature investigates to what extent professionals solve problems as scientifically knowledgeable practitioners (Chapter 2.6). If practitioners show difficulties with engaging in scientific reasoning, this might serve as an argument to investigate further whether implementing collaboration for problem solving may foster practitioners' scientific reasoning.

2.1. Scientific reasoning and problem solving

From early on there has been a line of research that understands scientific reasoning as a process of (heuristic) discovery and inference-making in order to solve a problem (Klahr & Dunbar, 1988; Simon et al., 1981; Zimmerman, 2000). Moreover, as such inquiry-based inference-making does not always have to be systematic and it can be analogous to general problem solving activity (Siegler, 1978; Simon, 1989). According to this "science-as-problem-solving view" (Klahr & Dunbar, 1989) scientific reasoning can be understood as a way of complex problem solving (Klahr & Simon, 1999) and, taken that view, scientific reasoning and problem solving skills are not considered as unrelated constructs (Zimmerman, 2000). What differences and what commonalities do they have then? As Simon et al. (1981) argues there are two distinct features that may be more the characteristics of scientific reasoning than of general problem solving, such as (1) the social nature and (2) the occasional goal-indefiniteness. The first one refers to that scientists often solve problems in collaboration over a longer period of time (Archer et al., 2010; Dunbar, 1995). The latter distinction refers to those occasions when the goals are not yet from the beginning of discovery clearly identifiable for scientists, compared to those typical problem solving scenarios where a clearly defined end-state exists. Besides these differences, the authors argue, scientific inquiry might consist of processes that "are not qualitatively distinct" from solving non-scientific problems. For those processes they propose data gathering, finding parsimonious descriptions and developing explanatory theories as well as the invention of new instruments or methods of observation. Theory development and contrasting it to evidence are indeed from early on and rather widely considered as central to both scientific reasoning and problem solving (Klahr & Dunbar, 1988; Kuhn, 1989; Sandoval, Sodian, Koerber, & Wong, 2014). As Kuhn (1989) argues in her review that while the skills of theory development, identification of relevant evidence and drawing appropriate conclusions based on that evidence might not capture all features of scientific reasoning, they are "the most central, essential and general skills" in reasoning scientifically (p. 674).

From this short theoretical introduction it seems reasonable to conclude that (1) general problem solving and scientific reasoning share similar features and that (2) some of these features, e.g., hypothesizing (Klahr & Dunbar, 1988) evaluating evidence (Kuhn, 1989) and engaging in a social negotiation process (Archer et al., 2010; Simon et al., 1981) might be more "core" to scientific reasoning (while these distinctive features can still occur in the context of general problem solving). The following subchapter focuses on the commonalities in order to point out the relevance of scientific reasoning for practitioners' problem solving, suggesting that the quality of practitioners' general problem solving might be indicated by the extent to which they engage in such processes that are also relevant or, as just mentioned, "core" epistemic processes for scientists (Fischer et al., 2014; Madsen & Olson, 2005; Sackett et al., 1996; Spencer et al., 2012).

2.2. The "scientifically knowledgeable practitioner"

It is arguable that a professional way to solve problems is to entertain a systematic process of reasoning, meaning to engage in certain epistemic processes that allow using and generating (profession-relevant) knowledge in order to solve a problem. Klahr and Simon (1999) note that in that respect scientists and practitioners "share the same general approach to solving their respective problems, and they use the same weak methods" (p. 540). By "weak" they refer to heuristic or cross-domain applicable methods rather than "not good enough" ways to solve a problem. Most certainly, an important and existing counter-argument is that professional problem-solving is not independent from specialized expertise in a given domain, meaning that the quality of problem solving would emerge from expertise in that domain instead of domain-independent problem solving heuristics (Anderson, 1987; Chi, Feltovich, & Glaser, 1981; Ericsson, 2006; Gilhooly, 1990; Hmelo-Silver, Marathe, & Liu, 2007). Yet, there are studies that suggest that expertise may not always have a positive effect on professional problem solving (e.g., Stark et al., 1998). Nevertheless, this dissertation is seeking to identify some domain-general features or even "heuristics" that, regardless of the domain, can be fruitfully applied, e.g., coordinating hypotheses with evidence (Klahr & Dunbar, 1988; Kuhn, 1989; Sandoval et al., 2014). In other words, following the idea of "weak methods", the present work proposes that certain epistemic processes might have relevance for solving problems in different practice-oriented domains, because they capture general epistemic, and therefore, rather domain-independent characteristics, of reasoning.

Following the earlier thought, Simon (1989), for example, specifies "typical" processes scientists engage in: e.g., forming problems or asking questions, drawing inferences, testing these inferences or inventing new instruments for observation. Most importantly, he adds that "What is common to all of these tasks is that they appear to employ the same general kinds of problem solving processes as are employed […] by physicians making diagnoses, by computer salesmen configuring systems for clients, by architects designing houses". This insightful observation supports the point of the present work, namely that the epistemic processes scientists engage in or the way practitioners solve problems might share epistemic commonalities. Further authors suggest that while solving problems practitioners, just like scientists, should be able to identify and analyze the problem at hand (Robinson, 1998; Weinstock, 2009), set up questions to investigate further (Hou, Sung, & Chang, 2009), generate assumptions or hypotheses about the problem (Charlin et al., 2007) develop artefacts (Dennen & Hoadley, 2013), rely on evidence (Spencer et al., 2012), evaluate that evidence (Morris, 2006) in order to come up with a conclusion (Rummel & Spada, 2005) and communicate the problem solving process across their peers (Nokes, Bullough Jr, Egan, Birrell, & Merrell Hansen, 2008).

Fischer et al. (2014) suggests a similar, yet, wider perspective when interpreting the relationship between scientific inquiry and practice-oriented problem solving by introducing the concept of "epistemic modes". According to their framework, scientific reasoning can follow different (epistemic) motives and, correspondingly, might occur in different epistemic modes. They propose three epistemic modes of scientific reasoning by building on the classification system of Stokes 1997 (in Fischer, 2014), and accordingly, identify a theory-building mode; an artefact-centered mode and a practice-oriented mode. The last one, the practice-oriented mode of scientific reasoning is characterized by solution development for problems that are contextualized in the professional practice. While solving the problem, the "scientifically knowledgeable practitioner" will, similarly to scientists, rely on "scientific concepts, theories and methods" (Fischer et al., 2014, p. 5). Finally, Fischer et al. (2014) makes a clear distinction between solving problems as an instance of scientific reasoning (or "science-based reasoning in practice", as they call, p. 6) and general problem solving is that while general problem solving has a main concern of finding a solution to the problem, scientific reasoning should additionally result in an "argument". Although at this point the authors (Fischer et al., 2014) do not explicitly specify what they mean under "argument" as a result, yet, based on their line of reasoning, they seem to understand that the science-based practitioner, compared to the general problem solver, 1) engages in processes of knowledge construction and at the same time 2) relies on scientific knowledge such as "concepts, theories and methods" (Fischer et al., 2014, p. 6).

To summarize, there are similarities between scientific reasoning and solving problems in practice regarding the processes that general problem solvers and scientists or "science-based practicioners" engage in (Klahr & Dunbar, 1989; Klahr & Simon, 1999; Simon, Klahr, & Kotovsky, 1989). Yet, beyond these heuristic similarities between scientific reasoning and problem solving, scientific reasoning might potentially be characterized by a greater extent of socially exchanging ideas (Rummel & Spada, 2005; Osborne, 2010; Simon et al., 1981), developing hypotheses (Charlin et al., 2007; Klahr & Dunbar, 1988), evaluating those hypotheses in light of evidence (Klahr & Dunbar, 1988; D. Kuhn, 1989; Simon et al., 1981) as well as using scientific knowledge in forms of e.g., theories and evidence (Fischer et al., 2014; Spencer et al., 2012). But is that all what scientific reasoning is? What else does scientific reasoning composed of? In the following, this dissertation is seeking to find a more conclusive answer to these questions than it has reached so far.

2.3. Scientific Reasoning

What is scientific reasoning? Although different interpretations are available on what constitutes scientific reasoning (Engelmann, Neuhaus, & Fischer, 2016; Fischer et al., 2014; Zimmerman, 2000), there is also a reasonable convergence in research as conceptualizations point to some identifiable directions as follows. Generally speaking, scientific reasoning is often understood as a complex skill or set of skills that affords knowledge construction in a systematic scientist-like manner (Chinn & Malhotra, 2002; Fischer et al., 2014) Osborne, 2010). As some authors (David Klahr & Dunbar, 1989; Zimmerman, 2000) point out, at least two main strands of research can be differentiated for investigating scientific reasoning: the abovementioned (1) problem solving approach and a (2) concept formation approach. As earlier discussed, the first one reflects on how reasoners solve problems (e.g., Okada & Simon, 1997; Rummel & Spada, 2005) while the second one reflects on the development in reasoners' understanding and use of different (e.g., scientific) concepts (e.g., Hmelo-Silver et al., 2007). Beyond these two main approaches, other authors (e.g., Chinn & Malhotra, 2002; Sandoval, 2005; Sandoval & Reiser, 2004) emphasize a further, an (3) epistemic dimension by arguing for the relevance of epistemologies in scientific reasoning (i.e., how scientific knowledge is created and used). This latter view is also close to a "nature of science" conceptualization of scientific reasoning (Engelmann et al., 2016, Koenig, Schen, & Bao, 2012)

Simon and Lea (1974) discussed an already existing division in the psychology literature between problem solving and concept attainment and made an attempt to bring these two areas of reasoning under a common framework. On the way they identified algorithms that can also model reasoning processes in this respect, including e.g., heuristic search. Continuing the attempt of reducing the division between the problem solving and the concept development approaches Klahr and Dunbar (1988) created an integrated framework for scientific reasoning. As one important conclusion of this attempt, Dunbar and Klahr (1989) stressed that both concept formation and problem solving tasks might involve similar cognitive or reasoning processes such as hypothesizing and experimenting. Similarly to these exemplary attempts, the interpretation of this work regarding scientific reasoning aims to connect the diverging research traditions and considers all the three approaches described above: the problem solving view, the conceptual approach and the epistemic view. First, this dissertation argues that driven by an epistemic aim to e.g., better understand a problem, one will engage in scientific reasoning processes (e.g., she will generate evidence in order to test a hypothesis). Second, this work considers that one’s reasoning can prove to show different epistemic qualities (e.g., considering the source of evidence to be used to test a hypothesis). More specifically, the present dissertation understands scientific reasoning as an epistemic pursuit of applying and gathering (scientific or everyday) knowledge in order to solve (scientific or everyday) problems. For the sake of conceptual clarity, however, in the following, this work keeps a distinction between the “problem solving”, i.e., the *process* aspect and the “knowledge application”, i.e., the *content* aspect of scientific reasoning. While it is argued that both of these aspects are relevant for scientific reasoning, the present work considers them as orthogonal, yet, complimentary dimensions that describe different qualities or features of scientific reasoning and should be simultaneously taken into account.

As a visual summary, Figure 1 depicts the theoretical conceptualization on scientific reasoning as introduced above. The following subchapter (2.4) explores further the process aspect while the subchapter after (2.5) will delve deeper into the content aspect of scientific reasoning.

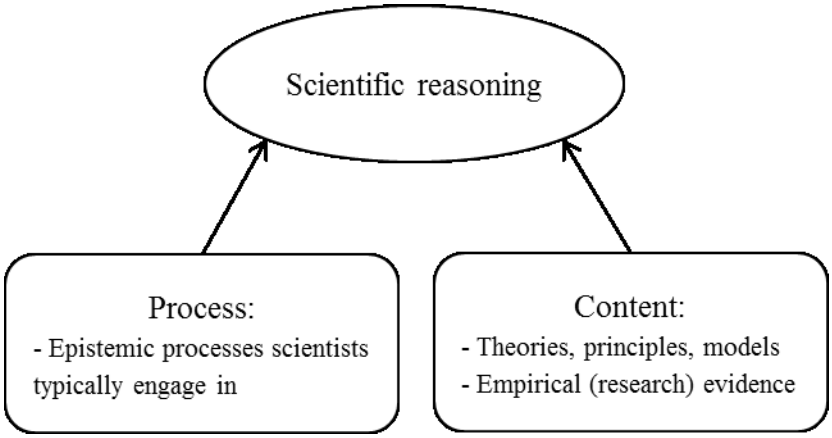

Figure 1. The perspective of this dissertation on scientific reasoning

2.4. Scientific reasoning processes

To follow the above introduced definition, it can be argued that regarding the process aspect scientific reasoning is an epistemic pursuit to solve a problem as well as to gather knowledge on the way which can support that problem solving process. This part of the conceptualization stems from the "science-as-problem-solving" and the inquiry-oriented approaches (Klahr & Dunbar, 1988; Dunbar & Klahr, 1989; Simon et al., 1989) as from this perspective this work views scientific reasoning as a systematic engagement in epistemic processes of scientific inquiry (Fischer et al., 2014). Along this theoretical and empirical tradition different epistemic processes have been identified to play a role in scientific reasoning and inquiry (de Jong, 2006; Fischer et al., 2014; White & Frederiksen, 1998). Simon (1989), for example, lists some processes scientists might typically engage in such as formulating problems, drawing inferences from theories, testing such inferences, designing experiments, explaining and generalizing outcomes, inventing new instruments for observation and so forth. These processes as well as further ones were also addressed by more recent research such as constructing a problem space (Weinberger & Fischer, 2006), questioning (Kuhn & Dean, 2005; White & Frederiksen, 1998), hypothesizing (Klahr & Dunbar, 1988; Lawson, 2002), generating evidence for hypothesis testing (Schauble, Glaser, Raghavan, & Reiner, 1991) and evaluating evidence (Klahr, Zimmerman, & Jirout, 2011; Sandoval et al., 2014). Fischer et al., (2014) proposed a framework for scientific reasoning and their model collects eight epistemic processes of scientific reasoning. The Fischer et al. (2014) model encompasses those processes that are reported across different theoretical or empirical strands of research, and in this respect it can serve as a useful framework for further (empirical) application. In the following, the present work introduces the model, i.e., the epistemic processes of scientific inquiry.

2.4.1. Problem identification

To start reasoning about a problem, the problem solver needs to realize that there *is* a problem to solve. As Fischer et al., (2014) describes: "During this epistemic activity, a problem representation is built from an analysis of the situation." (p. 5). There are similar operationalizations of problem identification. For example, (Weinberger & Fischer, 2006) argues for the epistemic process of "Construction of problem space" which is an early stage of problem solving when reasoners analyze the problem case "with the aim to foster understanding of the problem" (p. 74). Such initial understanding of the problem is essential to engage in further valid reasoning processes. Research on problem solving from early on has recognized that an initial step of the problem solving process is to build a (mental) representation or an understanding on what constitutes the problem (Chi, Feltovich & Glaser, 1981; Hinsley, Hayes & Simon; Simon & Newell, 1971). More recent studies on problem solving applying neuroimaging techniques also argue for the idea that there might be a distinguishable initial stage of problem solving (Anderson & Fincham, 2013; Chinn &

Malhotra, 2002; Kuhn & Dean, 2005; Simon et al., 1989; White & Frederiksen, 1998) which can be labeled as an "encoding" of the problem.

2.4.2. Questioning

As Fischer et al. (2014) proposes, based on the problem representation initial questions are developed that aim to move further reasoning processes forward. Besides problem identification, questioning is often emphasized as another critical initial step of scientific inquiry (Chinn & Malhotra, 2002; Kuhn & Dean, 2005; Simon et al., 1989; White & Frederiksen, 1998). White and Frederiksen (1998), for example, understands questioning as relevant in setting directions for further predictions that will eventually lead to better understanding of a topic of interest (which can be e.g., the problem itself). Kuhn and Dean (2005) argue that this stage of scientific inquiry can be critical for subsequent reasoning, as for example, "because it organizes and gives meaning to the activity that follows" (p. 869). For example, a teacher facing a problem of very heterogeneous performance of her students might ask herself why that is so. This will potentially lead her to develop some plausible explanations of that problem and she may collect further information (about the students or about the class) in order to draw more valid conclusions.

2.4.3. Hypothesis generation

Hypothesizing, developing investigable explanations or, in general, engaging in hypothetico-deductive reasoning is considered as a core epistemic process of scientific reasoning (Klahr & Dunbar, 1988; de Jong & Van Joolingen, 1998; Lawson, 2005; Okada & Simon, 1997; White & Frederiksen, 1998). During this epistemic process, following the specification of Fischer et al. (2014), reasoners "derive possible answers to the question" (p. 6). This means, that hypothesis generation can be viewed as a coherent part of the inquiry cycle stemming from earlier epistemic processes of problem identification and questioning while directing subsequent processes of evidence generation evaluation of this evidence in the relation of hypotheses. Such hypotheses can provide candidate explanations for and serve as a mean to better understand the original problem. For example, a physician listening to a patient's symptoms develop an immediate hypothesis about plausible illnesses which will also direct her further investigation in order to be able to set a proper diagnosis (Charlin et al., 2007). It may be important, however, to mention that the process of hypothesis generation may not only precede further information collection (i.e., evidence generation), therefore this "a priori" way of developing hypotheses is not exclusive for scientific reasoning. In addition to that, hypothesis generation can happen "a posteriori" as new hypotheses can be derived from evidence or even existing ones can be revised this way. Klahr & Dunbar (1988), for example, claims that hypotheses can be generated in two ways: based on prior knowledge as well as a generalization from evidence generation. Similarly, de Jong and van Joolingen (1998) argue for generating "a posteriori" hypotheses, i.e., that also disconfirming evidence should lead to hypotheses generation by revising one's initial (theoretical) assumptions. For instance, if a teacher initially believes that her pupil's performance is low due to lack of engagement, yet, the teacher learns that the pupil does invest in learning, then the teacher might want to revise her initial explanation of the problem and may try to develop new assumptions.

2.4.4. Constructing artefacts

According to Fischer et al. (2014) scientific reasoning often includes the development of a sort of artefact. The process of constructing an artefact is scientific in a sense that it incorporates scientific knowledge and is aimed to improve scientific reasoning (e.g., to alleviate the problem). This artefact may be physical or related to a physical system, for example, a teacher may want to construct a test following methodological and theoretical guidelines (Asim, Ekuri, & Eni, 2013; Gordon, Collins, & Jewell, 2015) in order to assess her students' understanding on a topic. This test as an exemplary artefact is optimally not only constructed by scientific knowledge (i.e., methodological and theoretical) principles, but it contributes to knowledge construction and engagement in further inquiry processes: it can provide evidence for evaluation that may lead to further hypothesis generation processes and so forth. Or to give another example, Fischer et al., (2014) mentions a prototype of a learning environment that follows the guidelines of theoretical design principles and potentially on earlier empirical evidence. However, such an artefact does not necessarily have to reach the stage where it is materialized. In fact, from the definition of Fischer et al. (2014), it might follow that it is the scientifically reasoned (i.e., principle-based) construction itself as well as the reasoning processes that the construction might lead to which can be considered as the scientific aspects of artefact construction and not the materialized outcome. Following the example of Fischer et al. (2014) on a learning environment based on design principles, it might matter less whether the teacher has managed to create e.g., an online learning environment. What matters is whether a plan of the environment is being constructed based on specific design principles (e.g., Dennen & Hoadley, 2013) and similarly, if the plan of such an environment allows further hypothetical or "what if" thinking processes even in case of the lack of available evidence (see e.g., conceptual simulation at Trickett & Trafton, 2007).

2.4.5. Generating evidence

Besides hypothesis generation, generating evidence is another epistemic process typically viewed as core to scientific reasoning (Klahr & Dunbar, 1988; Okada & Simon, 1997; Schauble et al., 1991).Yet, most of the aspects of generating evidence are indicated to be not expertise-dependent (e.g., planning and conducting systematic evidence collection; Schunn & Anderson, 1999) and therefore, this epistemic process can be considered also as a potential candidate to characterize general problem solving processes. Generating evidence, for example, can take form from everyday information seeking such as searching on the web (Goldman, Braasch, Wiley, Graesser, & Brodowinska, 2012; Lazonder, 2005) to controlled experimentation (Kuhn & Dean, 2005; White & Frederiksen, 1998). In a rather general way, it can mean to use or refer to any (scientific or everyday) information while reasoning (e.g., Hetmanek et al., 2015). For example, a teacher interested in how to increase her students' motivation might look for recent empirical findings on how to improve her methods in order to engage students. In this respect, evidence generation, the skill to search for, to find and to rely on evidence is a crucial process to work as an evidence-based or scientifically knowledgeable practitioner (Fischer et al., 2014; Hetmanek et al., 2015; Spencer et al., 2012).

2.4.6. Evaluating evidence

Evidence evaluation can be understood as the coordination between a hypothesis (or hypotheses) and the available evidence (Klahr et al., 2011; Kuhn, 2001, 2002; Sandoval et al., 2014). This epistemic process is mainly concerned with if the available evidence is in accordance with the initial assumptions (i.e. hypotheses) or if those assumptions should be revised due to the contradicting evidence. Another conception of evaluating evidence can be to assess the merits of the evidence-to-be-used. This form of evidence evaluation may involve some epistemic criteria regarding e.g., how evidence should be generated (Chinn & Malhotra, 2002; Sandoval et al., 2014). To give an example, a student may underperform in a subject but she may gain very good grades in other subjects. In this case, the teacher may be aware of the limitations that the evidence (student's low grade in one class) is generated under specific circumstances (e.g., a student underperforms only in that class but not in other classes). Consequently, the teacher would not use the evidence of underperformance to support an assumption of the students' general motivation or capabilities. This epistemic understanding about how to use evidence is especially relevant for scientifically knowledgeable practitioners (e.g., Koenig et al., 2012) in order to select appropriate (e.g., the “best available evidence”; Spencer et al., 2012) to support their decision-making and problem solving processes.

2.4.7. Drawing conclusions

Drawing conclusions is an epistemic process that has a focus on coordinating and assessing earlier inquiry processes and outcomes. Yet, it means more than a coordination process between two epistemic processes such as hypothesis generation and evidence evaluation which is in the focus of evidence evaluation (Sandoval et al., 2014). Drawing conclusions is a more complex epistemic process in a sense that it has a more conclusive assessment-focus than, for example, evidence evaluation. According to Fischer et al. (2014) it often includes the integration of different evidences by taking into account the methods by which these evidences were generated as well as the epistemic criteria of the discipline(s) involved so as to conclude something about e.g., the applicability of an artefact such as a learning environment or the characteristics of the original problem. For example, when summarizing a diagnosis collaborating professionals might evaluate the information they know about a patient in the light of reliable diagnostic criteria, judge the likelihood of certain types of illnesses and take into account potential therapy plans at the same time (e.g., Rummel & Spada, 2005). While doing so they consider (and may articulate) the outcomes of earlier epistemic processes.

2.4.8. Communicating and scrutinizing

Scientific reasoning often happens in a collaborative context (Sampson & Clark, 2008) (Osborne, 2010). As Simon et al., (1981) argues, one distinctive feature of scientific reasoning in comparison to problem solving is that “scientific inquiry is a social process, often involving many scientists and often extending over long periods of time” (p. 1). Communicating concerns or outcomes of the reasoning processes can be considered as a “wheel” to improve further inquiry at various stages. As Dunbar (1995) concludes his investigations on scientific collaboration, a key role of communicating and scrutinizing is that it may initiate "a chain of reasoning that can then result in a reconceptualization of a theory, data, or experimental design" (p. 14). For example, while exchanging knowledge, partners often develop deeper understanding of the original problem (Michelene T. H. Chi & Wylie, 2014; Roschelle & Teasley, 1995); generate alternative hypotheses (Okada & Simon, 1997), negotiate the interpretation of evidence (Osborne, 2010).

Practitioners also often need to engage in collaboration with each other, and often in an interdisciplinary way (Rummel & Spada, 2005). In such scenarios partners need to engage in productive talk in order to develop a similar representation of the problem, and to share all the necessary information in order to draw a valid conclusion. Studies from paired teaching (e.g., Nokes et al., 2008) suggest that although collaborating teaching practitioners may experience some difficulties e.g., when they need to coordinate their ideas to prepare for a class (Bullough et al., 2002); this negotiation between collaborating teachers is often beneficial as they learn from exchanging ideas (Baeten & Simons, 2014).

2.5. Scientific reasoning as applying scientific content

According to the definition of this work on scientific reasoning (see Chapter 2.3), reasoners can reason scientifically in two main aspects: (1) they can engage in epistemic processes of scientific reasoning (i.e., in epistemic processes as introduced above); and (2) they may apply knowledge acquired through scientific inquiry. The present work handles these two aspects as quasi-orthogonal in a sense that neither engaging in epistemic processes presupposes the use of scientific knowledge nor the use of scientific knowledge necessitates the engagement in epistemic processes.

To give an example for applying scientific content without engaging in epistemic processes of scientific reasoning, let us consider the following. A teacher may experience that her students are not engaged in class and their performance is also somewhat beyond her expectations. This teacher may attribute the problem to the students' lack of motivation. Accordingly, she may apply motivational theories in order to engage her students. At the same time, she is not conducting further inquiry (e.g., developing alternative hypotheses or evaluating evidence) in order to investigate the problem further and in order to derive more valid solutions for it. To be more concrete, the teacher might consciously implement strategies to raise her students' intrinsic motivation, e.g., via an increased sense of autonomy and competence (e.g., Ryan & Deci, 2000). Yet, this teacher, by not engaging in epistemic inquiry processes, may miss more valid explanations of the problem, for example, her students may have suboptimal prior knowledge and this impedes with their sense-making processes as well as it results in their lack of engagement. Consequently, the teacher's intervention may be less effective than one's performance could potentially be who is engaging in scientific reasoning processes. On the other hand, developing naïve explanations without considering scientific content knowledge can also lead to problems. To follow the before-mentioned exemplary problem, the teacher can engage in a systematic way of inquiry. As such, she can develop different hypotheses about the problem, she can contrast those to evidence and might even find the most plausible one: e.g., the students might have problems with understanding the material. Yet, without having a theoretical understanding of learning processes she might choose inadequate instructional techniques (e.g., she cannot scaffold the students according to the level of their prior knowledge; Fischer et al., 2013). For example, she asks questions or provides clarifications that are for students who have very different conceptual understanding on the learning material, and thus, this teacher may induce an unnecessary cognitively load on her students.

The examples mentioned above were intended to illustrate that both aspects of scientific reasoning (process and content) can work independently, yet, both might be necessary for working as a science-based practitioner and reaching justified conclusions from which valid solutions emerge. Therefore, in the following, this work continues the introduction with further clarification on the second aspect of scientific reasoning, i.e., the “knowledge application” or content aspect (see Chapter 2.3 and Figure 1) which can be considered, besides the process aspect, the other epistemic criterion for scientific reasoning.

This second aspect, the “knowledge application” aspect of scientific reasoning is only partially related to the conceptual development view on scientific reasoning (Zimmerman, 2000). As opposed to that view, the present work is less interested how reasoners develop their understanding on a scientific concept such as external motivation, but it is more interested whether reasoners are able to apply that concept during problem solving. In this sense, the content aspect of scientific reasoning is somewhat closer to the line of research on “evidence-based practice” (Hetmanek et al., 2015; Mandinach, 2012; Sackett et al., 1996; Satterfield et al., 2009; Spencer et al., 2012; Wenglein et al., 2015).

The question about what knowledge evidence-based practice should include might yield somewhat different answers across domains (see e.g., Spencer et al., 2012) and even across researchers (see e.g., Robinson, 1998). Yet, in the studies on evidence-based practice there seems to be an overarching agreement that the practice of applying empirical evidence accumulated through systematic observation (e.g., research evidence or field observation) should be a part of such evidence-based approach (Hetmanek et al., 2015; Satterfield et al., 2009; Spencer et al., 2012). According to this definition an evidence-based practitioner should always be objective and rely on empirically gathered information (e.g., Spencer et al., 2012).

The present work builds on the understanding of evidence-based practice, yet, to a limited extent. This dissertation claims that applying *any* type of knowledge created through *scientific* inquiry (i.e., research evidence) should be considered as scientific reasoning (see Figure 1). Or to be more specific, it would fulfill the content-criteria of reasoning scientifically. This definition, however, works with somewhat different borders than the definition of evidence-based practice. First, it *narrows* the evidence-based practice view in some respect. Specifically, the present work proposes that although empirical evidence that does not result from "*scientific* inquiry" (e.g., the teacher's observation of her students' behavior) should not be considered under the view of applying scientific knowledge as an instance of scientific reasoning. In this respect, the understanding of the present work on what constitutes "evidence" as knowledge to be applied is narrower than the proposals from evidence-based practice (Sackett et al., 1996; Spencer et al., 2012). In another respect, however, the definition of this work also *extends* the evidence-based approach. By claiming "*any* type of knowledge it is meant, that it is not only applied empirical (e.g., research evidence) but also applied theoretical knowledge (scientific theories and related models) originating from scientific inquiry should be considered as instances of scientific reasoning. To conclude, the present work argues that the "knowledge application" aspect, i.e., applying empirical or theoretical scientific knowledge is a relevant criterion for reasoning scientifically as a practitioner.

2.6. Do practitioners reason scientifically?

The previous subchapters elaborated on the conceptual definition of this dissertation on scientific reasoning claiming that "scientifically knowledgeable practitioners" (Fischer et al., 2014), similarly to scientists, also engage in systematic reasoning processes (Klahr & Simon, 1999) applying scientific knowledge (Spencer et al., 2012) while solving practical profession-related problems (Rummel & Spada, 2005) However, to what extent does empirical evidence support that notion? To be more specific, are future practitioners as novice problem solvers able to demonstrate scientific reasoning skills (as a problem solving heuristics) so as they can be considered as “scientifically knowledgeable practitioners”?

As different sources of evidence seem to show, it may not always be the case. In their review, de Jong (2006) found that learners often have problems with hypothesis generation, evidence generation-evaluation and drawing conclusions. The authors’ main concern is that students are not systematic enough while engaging in such processes. For example, they do not use control of variable strategy, i.e., to change only one variable at a time so as to be able test its effect. Also, students often fall to confirmation bias, i.e., they seek for evidence that “supports” their initial hypotheses and argumentation, and therefore, miss or neglect evidence that contradicts to their assumptions. Clearly, such cognitive pitfall is in direct contradiction with the expectations of being evidence-based and to find the best available evidence regardless of motivational biases of the reasoner (e.g., Spencer et al., 2012).

Other studies further confirm the notion that coordinating hypotheses with evidence (e.g., making evidence-based decisions or revising the initial assumptions based on new evidence) is often problematic for university students (Klahr & Dunbar, 1988; Stark et al., 2009) and even for teaching professionals (Chinn & Brewer, 1993; Yeh & Santagata, 2015; van de Pol, Volman, & Beishuizen, 2011). Schunn and Andersson, (1999) for example, mentions that university students often do not use theories when planning evidence generation and they often rely on personal beliefs instead of evidence when drawing conclusions. In addition, Stark et al., (2009) reports that students often misinterpret evidence by not appropriately identifying the effect of a variable on another: e.g., they attribute the effect to another variable or they incoherently switch between different explanations. Such misinterpretation of evidence and drawing invalid conclusions is also reported by studies that are concerned with teachers' diagnostic errors (e.g., Wittwer, Nückles, & Renkl, 2010). While future teaching professionals need to estimate their students' knowledge in order to plan their instructions accordingly (Pickthorne, 1983; Pol, Volman, Oort, & Beishuizen, 2014), empirical research shows that (future) teachers have difficulties in drawing valid evidence-based conclusions (Chi, Siler, & Jeong, 2004; Morrison & Lederman, 2003; Yeh & Santagata, 2015). Moreover, teachers' invalid assumptions regarding their students' performance as well as the difficulties teachers show when evaluating evidence in order to reconsider their initial assumptions, can lead to premature solutions to the educational problems teachers need to solve in their practice (e.g., Chi et al., 2004).

Besides the abovementioned, rather process-related challenges for future practitioners' scientific reasoning, there are also content-related inaccuracies that often occur. For example, Stark et al. (2009) mentions that students often refer to own experiences instead of scientific theories or empirical findings and they show conceptual difficulties with scientific theories (e.g., by referring to the wrong theory or mistakenly referring to a theory). Similarly, other studies (Schunn & Anderson, 1999) also report it as an often upcoming problem that university students refer to personal opinion or "anecdotal" evidence instead of scientific theories or evidence. Further studies with teaching practitioners (Hetmanek et al., 2015) report similar issues, such as rarely using scientific resources when solving professional problems. In accordance with these findings, others (Morrison & Lederman, 2003) found that when teachers diagnose students' understanding (e.g., preconceptions), they do not refer to scientific content such as "concept maps" or "writing prompts" (the examples are from the authors). There are other studies indicating that teachers may not appropriately consider theoretical knowledge from their university training when making instructional decisions

(McElvany et al., 2012). Further research (Gruber et al., 2000; Stark et al., 1998) point to the problem of transferring knowledge from research to practice when solving complex professional problems.

All in all, from the empirical studies it is becoming visible that future practitioners often do not meet the criteria of a science-based practitioner (Fischer et al., 2014). They seem to have difficulties with engaging in scientific reasoning regarding both aspects of it: i.e., the process and the content aspects. Therefore, it seems reasonable to plan an “intervention” for fostering the engagement of future practitioners in scientific reasoning. Specifically, since scientific reasoning is arguably a social phenomenon, the present work investigates whether collaboration as a social context can be fruitfully applied to engage future practitioners in scientific reasoning, and therefore, to use collaboration in higher education to facilitate the acquisition of scientific reasoning skills. The assumption of this dissertation is that collaboration should be a way of educational intervention to engage in future practitioners in scientific reasoning. In order to see if such an assumption can be held to guide further empirical work; the upcoming chapter investigates to what extent this presumption can be justified. Therefore, the following chapter reviews the research on collaborative vs individual scientific reasoning.

3. Collaborative (vs. Individual) Scientific Reasoning

The present work argues that a collaborative context might facilitate both the process (i.e., engagement in epistemic processes) and content (application of scientific knowledge) aspects of scientific reasoning during practical problem solving. There are both theoretical and empirical reasons for this assumption and in the following this work takes a more detailed look at those in order to investigate why scientific reasoning should benefit from a collaborative setting.

First, for scientific reasoning collaboration may be a more *authentic context* than individual reasoning. Mercier & Sperber (2011) for example, claim that reasoning have evolved through communication with others and, consequently, gained an argumentative function to improve (both quantities and qualities of knowledge exchange across reasoners. In this sense reasoning has the social relevance of mediating information sharing process in groups and communities. Simon et al. (1981) while drawing a thin line of distinction between scientific reasoning and problem solving mention that one important difference between scientific reasoning and general problem solving is that "scientific inquiry is a social process" (Simon et al., 1981, p. 1). Similarly, Osborne (2010) argues that engaging in collaborative discourse and argumentation is a critical tool to develop scientific reasoning skills while it also facilitates conceptual understanding. Similarly, approaches of social constructivism emphasize the relevance of social interactions in reasoning and knowledge construction (see e.g., Brown, Collins, & Duguid, 1989; Lave & Wenger, 1991/2002; Pea, 1993). This can mean that a professional-to-be can only engage in an optimal level of reasoning if she is engaging in collaboration in order to work on problems that are authentic for her future practice. Such engagement in authentic practices (and collaborative inquiry might be considered as such for scientific reasoning) is also important to develop an epistemological understanding about the processes and criteria of knowledge-building within a community (Chinn & Malhotra, 2002; Sampson & Clark, 2008).

Second, collaboration has *communicational affordances* (e.g., engaging in interactive discussions; Chi, 2009) that are claimed to be advantageous for scientific reasoning (Dunbar, 1995; Osborne, 2010) as well as to general problem solving (Barron, 2000; Kirschner, Paas, Kirschner, & Janssen, 2011). Reasoning partners in a collaborative situation have the advantage to take into each other's ideas and build their reasoning processes on each other's contribution. Early research on collaborative vs. individual reasoning demonstrated how engaging in discourse can advance reasoning and problem solving (Berkowitz & Gibbs, 1983; Hill, 1982; Teasley, 1995) (Among the suggested mechanisms which can contribute to the advantage of collaborative reasoning, elaboration on one's own contribution as well as on the partners' contribution is typically identified (Kruger & Tomasello, 1986) (Berkowitz & Gibbs, 1983; Teasley, 1997). Such contributions can be part of a coordination process (Barron, 2000) where the aim of reasoners may be to develop and maintain a common understanding between each other (Clark & Brennan, 1991; Clark & Schaefer, 1989; Roschelle & Teasley, 1995). Furthermore, interactive processes (Chi, 2009), e.g., elaborating on the partners' ideas, are also characteristics of (collaborative) scientific discovery (Dunbar, 1995; Okada & Simon, 1997). As Dunbar (1995) summarizes his observations about collaborative scientific work in laboratories, he points out the importance of such interactive reasoning for scientific inquiry. He mentions that scientists often face surprising results or other problems that engages them in interactive discussions which, as a result, often leads to a complete re-evaluation of the inquiry process: "often one person's reasoning became the input to another person's reasoning. This resulted in a rapid reconceptualization of problems and to significant changes in all aspects of the way the research was conducted."

A third advantage of collaboration for scientific reasoning processes can be a cognitive advantage, i.e., the *distribution of cognitive resources* (Hollingshead & Brandon, 2003; Hutchins, 1995; Pea, 1993). Such advantage can be important, because reasoning scientifically during solving problems might be cognitively demanding. Although some studies indicate that some reasoning processes, for example, problem identification, can happen even rather quickly and effortlessly (Anderson & Fincham, 2014; Hinsley et al., 1977) unless further sense-making processes (Roschelle & Teasley, 1995) are involved. Yet, further studies (Kirschner, Sweller, & Clark, 2006; Shehab & Nussbaum, 2015) demonstrates that inquiry processes can often be cognitively engaging beyond such a "quick and effortless" categorization process and novices often suffer from such cognitive load when engaging in solving problems based on inquiry (e.g., Kirschner et al., 2011; Sweller, 1988). On the other hand, studies (Hinsz, Tindale, & Vollrath, 1997; Kirschner, Paas, & Kirschner, 2009; Kirschner et al., 2011) propose that collaboration might be even beneficial for problem solving, because the group can function as an information processing system and as such there is a lower cognitive load imposed on its individual members. Collaborative reasoners can benefit from forming such cognitive system, because individual members can rely on each other's cognitive resources distributing knowledge across each other (Hollingshead & Brandon, 2003; Noroozi et al., 2013) when engaging in potentially demanding tasks such as solving complex problems (e.g,, Kirschner et al., 2011).

While further empirical studies often support the notion that groups might perform better than average individuals do when engaging in problem solving and inquiry (e.g., Lazonder, 2005; Nokes-Malach, Richey, & Gadgil, 2015), other studies (Métrailler, Reijnen, Kneser, & Opwis, 2008; Sampson & Clark, 2009) also indicate that the results are not necessarily unequivocal. Additional studies draw the attention of further potential limitations of group performance. These limitations influence whether collaboration can be brought to reach its potential (Kirschner et al., 2009; Nokes-Malach et al., 2015; Paulus, 2000). Thus, in order to engage in productive collaboration, reasoning partners often need to overcome certain difficulties (Rummel & Spada, 2005) such as communicational difficulties to establish and maintain a mutual understanding of the problem they are working on (Clark & Brennan, 1991; Roschelle & Teasley, 1995). Another problem might occur if collaborators converge their ideas too fast, and by doing so, they obstruct knowledge exchange. This may happen when group members excessively rely on what everyone seems to know in a group instead of discussing knowledge in depth or considering alternative, unique knowledge (Stasser & Titus, 1985). Further factors (e.g., production blocking, cognitive load, social loafing, group size or evaluation apprehension etc. have been identified to account for such process losses moderating collaborative performance. Also, reviews (Nokes-Malach, Richey & Gadgil, 2015) and meta-analyses (e.g., Vogel, Wecker, Kollar & Fischer, 2016) underscore the notion that groups might need further (e.g., instructional) support to reach their potential.

Still, besides the limitations introduced above, considering that the main concern of the present work is (a) scientific reasoning situated in (b) problem solving, the abovementioned three arguments (authentic context, communicational affordances, distributed cognition) seem to be robust enough to assume that establishing collaboration would be advantageous for future professionals and this advantage can be realized through both (a) scientific reasoning (Okada & Simon, 1997; Osborne, 2010; Simon et al., 1981; Teasley, 1995) and problem solving (Hinsz et al., 1997; Kirschner et al., 2011).

4. Group Composition Problem

Nevertheless, problem solving groups can be very different from another regarding their members' demographic background (Curşeu & Pluut, 2013; Wegge, Roth, Neubach, Schmidt, & Kanfer, 2008); expertise (Noroozi et al., 2013; Wu, Liao, & Dai, 2015); prior knowledge (e.g., collaboration scripts; (Fischer et al., 2013) and so forth. This heterogeneity of group composition can have a significant impact on how groups collaborate, perform and jointly solve problems (Bowers et al., 2000; Canham, Wiley, & Mayer, 2012; Noroozi et al., 2013; Rummel & Spada, 2005).

Indeed, future practitioners have diverse backgrounds and, in addition, they might show very different approaches on solving complex problems and whether and how they would engage in scientific reasoning while solving problems. For such individual problem solving approaches the term "problem-solving scripts" will be applied following the "script" terminology of Schank (1999) as well as its more recent interpretations e.g., on "collaboration scripts" (Fischer, Kollar, Stegmann & Wecker, 2013). In that respect, scripts or "internal scripts" (Kollar, Fischer, & Slotta, 2007) represent a form of procedural knowledge that guide a person's understanding of and behavior in certain situations (Schank, 1999). Such scripts can contain knowledge about different aspects of a situation, most relevantly they tell what processes and in what possible sequence one might perform when entertaining a task such as solving a problem (Fischer et al., 2013). To be more specific here, problem solving scripts can be understood as a sort of "epistemic scripts" (Weinberger, Ertl, Fischer, & Mandl, 2005) as they represent an individuals' knowledge about in what epistemic processes (e.g., hypothesis generation, evidence evaluation) and in what sequence (e.g., generating a hypothesis before evaluating evidence) to engage in so as to solve a problem. Furthermore, it is argued that the extent to which these scripts consist of epistemic processes relevant to scientific reasoning

(Fischer et al., 2014) the more scientific they can be considered. This means that a person may understand problem solving completely in terms of epistemic processes of scientific inquiry while another person may see the very same problem as completely "practical" in a sense that it has nothing to do with engagement in scientific reasoning processes. Or to use an analogy, similarly to the expert/novice dimension on problem solving "strategies"(Chi, Feltovich, & Glaser, 1981; Ericsson, 2006; Voss, Greene, Post, & Penner, 1983), the present work argues that there might be a "scientific/non-scientific" dimension of problem solving that can show individual differences and thus, might be fruitfully used for creating heterogeneous/homogeneous groups regarding their prior knowledge (i.e., problem solving scripts) beyond e.g., the expert-novice dimension (e.g., Wiley & Jolly, 2003).

According to the Script Theory of Guidance (Fischer et al., 2013) such scripts (e.g., on problem solving) can have an effect on the groups' performance by guiding collaborating partners' engagement in collaborative processes (e.g., scientific reasoning). More specifically, the more divergent strategies reasoning partners have, the more they may stimulate each other during solving a problem, or even the opposite: the more conflicts they may face.

One argument is that reasoners solving problems in heterogeneous groups might stimulate each other (Dunbar, 1995; Paulus, 2000), e.g., by exchanging their different perspectives on a problem solving task (Miura & Hida, 2004; Stroebe & Diehl, 1994). Similarly, reasoning partners' diverse problem solving scripts might be complementary (Paulus, 2000) to each other leading to a scaffolding effect (Fischer et al., 2013) on and even to conceptual change through e.g., filling the gaps in each other's knowledge (Chi, 2008). Empirical studies (e.g., a meta-analysis by Bowers et al., 2000) indicate that such often assumed advantages of group diversity are likely to be a matter of task-type. The authors (Bowers et al., 2000) conclude that on the one hand, homogeneous groups seem to perform better on tasks that are clearly defined and require simple responses from the problem solver (e.g., solving a puzzle). On the other hand, heterogeneous groups seem to perform better on complex problem solving tasks that require cognitive engagement (e.g., business games). Further empirical studies (Canham et al., 2012; Wiley & Jolly, 2003) support their findings that group heterogeneity is beneficial for innovative problem solving that resemble more on complex than easy-to-solve tasks. Regarding scientific (analogical) reasoning, Dunbar (1995), for example notes that “when a problem arises, a group of similar minded individuals will not provide more information to make analogies than a single individual” (p. 13). Later he adds “Members of a research group should have different, but overlapping research backgrounds. This will foster group problem solving and analogical reasoning.” (p. 16). These notions are in accordance with the empirical findings on group problem solving suggesting that complex tasks such as solving authentic problems as a scientifically knowledgeable practitioner might benefit from group heterogeneity.

Yet, it is important to consider those empirical observations which emphasize that often heterogeneous groups may need more effort (e.g., time, engagement) to develop a common understanding of the problem and to coordinate their strategies to solve it (Bullough et al., 2002; Rummel & Spada, 2005). In addition, studies (see e.g., Hirst & Echterhoff, 2012) for a review demonstrated that collaborating partners having different memory retrieval strategies might face the phenomenon of collaborative inhibition through retrieval disruption, i.e., they recall less items than they would individually do due to the different retrieval strategies they apply. Such results indicate that heterogeneous problem solving scripts might similarly result in a collaborative inhibition effect. Namely, group members’ different problem solving strategies can not only complement but also interfere with each other and in this way they may hinder the activation of some of the individual members’ prior knowledge.

For example, when collaborating teachers leading a class together have different approaches to engage their students, it might lead to coordinating problems: e.g., one of them may want to find out whether the students are not interested or maybe have some misunderstanding, while the other teacher would disagree and implement a solution as quickly as possible. It is possible that ideas for alternative explanations on the problem may remain hidden due to such disagreements between collaborating partners.

All in all, the effects of group heterogeneity regarding group members' problem-solving scripts on the groups' engagement in scientific reasoning processes has not yet been empirically addressed. Considering the abovementioned studies, it seems reasonable to conclude that groups might benefit from diverse thinking as long as they have some shared knowledge based on which they can establish a common understanding that guides them while moving on with solving the problem.

5. Measurement of Scientific Reasoning

A second issue to consider regarding the main question of this dissertation, i.e., comparing collaborative vs individual problem solving with respect to scientific reasoning, is the measurement of scientific reasoning. As earlier (see Chapter 2) proposed, scientific reasoning can be understood as a complex process with two main components: process and content aspects. The process aspect was introduced through different epistemic processes in which a scientifically knowledgeable problem solver can be expected to engage in. Meanwhile, the content aspect was defined as the application of scientific (e.g., empirical, theoretical) knowledge when solving problems. In this two-dimensional approach, while the content aspect seems to be specific enough for further operationalization, analysis and conclusions; the process aspect, due to its multiple components (i.e., the eight epistemic processes), leaves probably more degrees of freedom for operationalization, analysis and conclusions than optimal. To be more specific, on the one hand, the content aspect, following its earlier definition (Chapter 2.5), can be operationalized and measured in a relatively problem-free manner by applying quantitative content analysis (Rourke, Anderson, Garrison, & Archer, 2000; Chi, 1997): e.g., by measuring the extent to which participants refer to scientific content while solving a problem ("coding and counting" strategy). Here, the conclusion would also come more or less clear: e.g., if groups refer on average more to scientific content than individuals do, it could indicate that collaborative problem solving is more science-based than individual problem solving. Therefore, it seems that the unidimensional content aspect of scientific reasoning can be captured by a "coding and counting" approach of content analysis.

However, the same "coding and counting" approach for measuring the process aspect of scientific reasoning (i.e., as engagement in the eight epistemic processes) would lead to difficulties regarding how to operationalize and measure scientific reasoning as well as what conclusions one can make about scientific reasoning after the analysis. To give an example, the process aspect of scientific reasoning can be indicated by the engagement in each of the epistemic processes. If a researcher decides to apply the "coding and counting" approach to measure scientific reasoning, and the results show that e.g., groups engage more in hypothesizing but less in evaluating evidence compared to individuals, then what could be concluded from that with respect to scientific reasoning? In that case the researcher faces a dilemma whether collaborative or individual problem solving could be considered as more science-based. This dilemma could be in part resolved if the researcher rephrases the question, and asks *in what respect* could collaborative problem solving be considered as more science-based than individual problem solving. Yet, the example shows that while the "coding and counting" approach allows the researcher to inspect how collaboration may affect each component (i.e., epistemic processes) of scientific reasoning; the main limitation of this approach is that conclusive statements regarding the overall scientific reasoning are difficult to be made.

On the other hand, an alternative methodological framework that affords the measurement of scientific reasoning on a "higher level" (i.e., overall scientific reasoning throughout the discourse) derived from its elementary components (i.e., epistemic processes) could solve the abovementioned problem. Measuring scientific reasoning on the "higher level" could inform the researcher about the overall "scientific quality" of problem solving. In this way, it could afford more robust comparisons between reasoning contexts (e.g., collaborative vs individual reasoning). For example, although studies indicate that collaborative reasoners compared to individuals engage to a different extent in some epistemic processes such as hypothesis generation and evidence evaluation (Okada & Simon, 1997; Teasley, 1995); there is a lack of scientific knowledge about whether there are indicative reasoning patterns for collaborative vs individual problem solving. Identifying such patterns as “higher level” indicators of scientific reasoning could allow further analyses, e.g., on the impact of different (collaborative vs individual) patterns of scientific reasoning on problem solving outcomes and learning gains (Chi, 2009; Suthers, 2005). Yet, the question is how to measure scientific reasoning on a “higher level” that accounts for all the elementary components, i.e., epistemic processes, involved.

To answer that question, i.e. how to derive information about the overall scientific reasoning from its components, it is important to clarify the (theoretical) relationships of those components to each other that might be also indicative to the subsequent methodology: i.e., their non-orthogonality. This dissertation argues that epistemic processes are not independent from each other. They probably reflect on different interrelated stages of scientific inquiry (e.g., Fischer et al., 2014). For example, questioning might be directly related to evidence generation (Kuhn & Dean, 2005); hypotheses might be derived from available evidence (Klahr & Dunbar, 1988); while, on the other hand, evidence is optimally evaluated in the context of one or more hypotheses (de Jong & Van Joolingen, 1998; Kuhn, 1989; Osborne, 2010; Sandoval et al., 2014). Thus, this dissertation understands scientific reasoning as a coherent process where epistemic processes evolve in relation to each other rather than occur as independent constructs. Therefore, the measurement of scientific reasoning on a “higher level” should also take into account the inter-relatedness of epistemic processes.

Beyond the abovementioned problems, focusing on isolated, independently measured processes without an accompanying analysis of reasoning process on the “high level” (e.g., by identifying patterns of reasoning) is an existing methodological problem also in the literature (Chi, 1997; Jeong, 2005; Klahr & Dunbar, 1988; Suthers, 2005). As, for example, Klahr and Dunbar (1988) notes: “the most implicit strategy in most psychological studies of scientific reasoning has been to investigate each skill in isolation […] much remains to be learned about how the stages interact and about how the interaction is influenced by prior knowledge" (p. 2). Indeed, many authors emphasize that scientific reasoning is a complex process and that reasoners may engage in a series of epistemic processes during scientific reasoning (Chinn & Malhotra, 2002; Fischer et al., 2014; Osborne, 2010; White & Frederiksen, 1998; Zimmerman, 2000). Yet, "coding and counting" approaches seem to dominate the analysis of scientific reasoning (Klahr & Dunbar, 1988; Métrailler et al., 2008; Okada & Simon, 1997; Simon et al., 1981; Teasley, 1995).

To summarize, analyzing the process aspect of scientific reasoning by applying a merely "coding and counting" strategy leads to methodological concerns. These problems, as mentioned above, seem to necessitate the search for a methodological formalism (Chi, 1997) which affords to (1) identify how elementary (epistemic) processes of scientific reasoning relate to each other (e.g., how they co-occur over time during the discourse); and to (2) draw valid conclusions about the overall process of scientific reasoning beyond the engagement in individual epistemic processes. In the following, this dissertation inspects whether reasoning patterns stemming from the relations between epistemic processes might serve as "higher level" indicators of scientific reasoning that are capable to fulfill the here mentioned criteria. To be more specific, the present work investigates which analytical approach, i.e., sequential analysis or network analysis, might be more appropriate to identify such patterns of scientific reasoning.

5.1. Sequential analysis

One methodological approach to identify patterns between epistemic processes of scientific reasoning could be sequential analysis (Bakeman & Gottman, 1997; Chiu, 2005; Cress & Hesse, 2013; Jeong, 2005). Sequential analysis goes beyond the "coding and counting" approaches in the following way. While a "coding and counting" approach is a useful method to identify and summarize occurrences of e.g., epistemic processes within a discourse; sequential analysis can be used to identify in what sequence those epistemic processes "typically" follow each other in the discourse (e.g., Cress & Hesse, 2013). Researchers assess this "typicality" by calculating the *transitional probability*, i.e., the probability that one process follows another (Jeong, 2005; Cress & Hesse, 2013). Once such transitional probabilities between epistemic processes are quantified (e.g., as odds-ratio or Yule's Q: see e.g., in Eells et al., 2011) the researcher can statistically compare (e.g., by applying z-scores; (Jeong, 2005) whether the probability of one sequence is higher than the probability of another sequence. Moreover, patterns such as chains of reasoning, e.g., transition diagrams (Cress & Hesse, 2013; Erkens, Kanselaar, Prangsma, & Jaspers, 2003; Jeong, 2005) can be identified, where the strength of connection between epistemic processes can be represented in the model (see Figure 3).

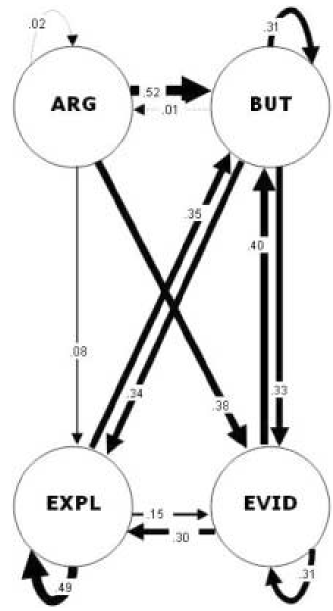

Figure 3. Transition probability pattern. Source: Jeong (2005).

Sequential analysis as well as its alternatives (Chiu, 2005; Chiu, 2008; Zhao, Sinha, Black, & Cassell, 2016) has been applied with success for analyzing reasoning processes during problem solving in practical professions such as psychotherapy (Eells et al., 2011), business (Jeong, 2003) and teaching (Hou et al., 2009). Yet, while this approach is useful to analyze sequential patterns in data (Chiu & Khoo, 2005), it yields patterns of predictive (Jeong, 2005) relations between epistemic processes. For this reason, sequential analysis can be warranted best if the researcher has a good (e.g., theoretical) reason to assume that predictive relations can reliably model discourse processes coded in the data or in cases when the question of interest is predictive. Nevertheless, there may be exploratory purposes to identify transition probability patterns such as the one demonstrated by Figure 3. In that case, arrows representing predictive relations (transitions) can show in both directions. However, the present work argues that in case a predictive (or unidirectional) assumption cannot be supported, measurement approaches that deal with non-directional relationships (e.g., co-occurrences) between epistemic processes can be more parsimonious, and consequently, more reasonable choices.

Furthermore, although analyzing process sequences (Cress & Hesse, 2013; Jeong, 2005) can fulfill the earlier mentioned criteria for a methodological approach, i.e., being able to capture relations between epistemic processes; it may be less powerful when it comes to the other earlier-set criteria: namely affording drawing conclusions about the overall reasoning process. On the one hand, transition probability patterns (see Figure 3) can represent the overall reasoning process, and “typical” sequential patterns can be extracted by applying this approach (Cress & Hesse, 2013; Erkens et al., 2003). On the other hand, a “summary” of the model seems to be problematic. For example, it can be problematic to extract a quantification that can represent the overall transition pattern and that also allows a comparison of transition patterns in that respect. Therefore, to find a metrics that can quantitatively represent the overall pattern of reasoning is necessary in order to compare collaborative and individual reasoners.

5.2. Network analysis

An alternative approach to sequential analysis in order to (1) identify connections between reasoning processes as well as to (2) be able to make conclusive statements regarding the overall reasoning in a discourse might be to investigate if reasoning processes may form a network, i.e., a pattern of interconnected processes. The idea of network identification for representing knowledge is based on semantic network models that were developed in the late 1960s (Collins & Quillian, 1969). Those early frameworks attempted to model memory subsystems, i.e., semantic memory by demonstrating how different components of knowledge are interconnected with each other. Their main proposal was that (semantic) knowledge components are organized in a network structure that can be described by its elementary units such as nodes and connections between the nodes, where nodes would represent concepts and their connections represent how strongly those concepts are associated with each other and, therefore, predict the accessibility of one concept in the presence of another (i.e., the likelihood and speed of memory retrieval; Anderson, 1983; Collins & Loftus, 1975). Further research on conceptual networks (e.g. Chi & Koeske, 1983) used interview data in order to identify to what extent elaborated knowledge (i.e., the density of conceptual networks) may predict later performance.

Beyond identifying associations, i.e., networks in conceptual knowledge, more recent studies pose the question how knowledge can be shared among collaborating partners, i.e., within a social network (Aviv, Erlich, Ravid, & Geva, 2003; Borgatti, Mehra, Brass, & Labianca, 2009; de Laat, Lally, Lipponen, & Simons, 2007; Oshima, Oshima, & Matsuzawa, 2012; Palonen & Hakkarainen, 2000). Those studies look at social network analysis as a useful toolkit of learning analytics (Aviv et al., 2003; Ferguson, 2012; Palonen & Hakkarainen, 2000) that can identify networks of interactions between learning partners in order to find out how knowledge sharing processes could be optimized.

From the approaches of conceptual networks and social network analysis it seems that network analysis may be an applicable methodological approach to analyze epistemic processes in verbal data. Also, networks provide quantifiable features (e.g., density of connections; (Chi & Koeske, 1983) that can be useful for analyzing how epistemic processes might relate to each other throughout the overall discourse. Such a quantified representation of overall patterns of scientific reasoning can make network analysis a more applicable methodological approach compared to sequential analysis when contrasting collaborating partners with individuals regarding scientific reasoning. Furthermore, network analysis might be also preferable over sequential analysis when the researcher does not have a (theoretical) justification to interpret the connections between epistemic processes in a sequential (i.e., predictive) manner.

All in all, network analysis seems to be an optimal methodological framework to identify patterns of scientific reasoning, because it fulfills both criteria set earlier in this chapter for a methodological framework. Namely, it affords (1) the identification of relationships between epistemic processes and (2) the interpretation of those relations on a higher level, i.e., drawing conclusions regarding the overall scientific reasoning process in the discourse. Yet, the question is what concrete methodology to apply to identify patterns of relationships between epistemic processes of scientific reasoning. Therefore, in the next chapter (Chapter 6) this dissertation introduces a specific approach of network analysis that affords the analysis of epistemic processes and provides additional features beyond traditional network analysis techniques (Shaffer, Collier, & Ruis, 2016). These features can support the analysis on the way of answering the main research question of this dissertation, namely comparing collaborative and individual scientific reasoning.

5.3. Epistemic network analysis

A possible approach to analyze patterns (i.e., networks) of scientific reasoning can be epistemic network analysis (ENA; Shaffer et al., 2009) ENA was originally developed to assess professional skill acquisition through the changes in epistemologies of professionals-to-be (e.g., (e.g., Hatfield, 2015). The method analyzes discourse patterns by looking at co-occurrences between codes. Based on the observed co-occurrences epistemic networks can be identified. Also, ENA offers quantification as well as qualitative representations of those networks for further analysis (Shaffer et al., 2016). In the following, the present work introduces this method in more detail to inspect how this type of network analysis might be a useful methodological tool to extract patterns of scientific reasoning from process (i.e., discourse or think aloud) data in order to use it for further quantitative and qualitative analyses.

ENA is a method to identify epistemic networks based on co-occurrences between discourse processes (Collier, Ruis & Shaffer, 2016; Shaffer et al., 2009). As Shaffer et al. (2016) points out, while many network analysis methods can measure certain features of networks (e.g., density), they are often optimized to deal with different sort of data (such as larger datasets with a high number of nodes); and provide summary statistics (e.g., density) that are not that satisfactory for datasets with relatively small number of nodes and often dense connections of weighted connections. At the same time, as the authors clarify, ENA can deal with networks containing a limited number of nodes and a large number of dynamic connections. This can apply for identifying a limited number of epistemic processes in verbal data.

The methodological approach of ENA (1) identifies co-occurrences between coded discourse processes by the the method of “moving stanza window” (Siebert-Evenstone et al., 2016): and (2) provides qualitative as well as quantitative “summaries” of the observed co-occurrences between discourses: e.g., “average” epistemic networks of a group of individual reasoners compared to the “average” epistemic networks of the group of collaborative reasoners (Shaffer et al., 2016).

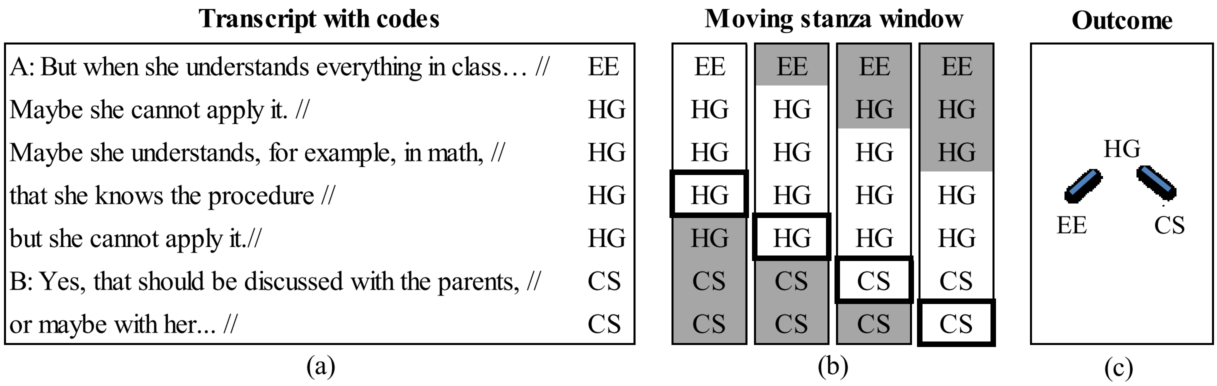

Figure 4. Illustration of the “moving stanza window” concept of ENA: (a) Exemplary coded transcript; (b) Moving reference unit (bold-framed) with moving stanza window; (c) The observed co-occurrences between epistemic processes (based on Siebert-Evenstone et al., 2016).

Figure 4 illustrates the concept of the “moving stanza window” method. This method allows the observation of co-occurrences between a selected “reference unit” (see bold-framed codes on Figure 4b) and a number of preceding codes which codes together constitute the “stanza window” (see the three codes preceding the bold-framed codes on Figure 4b). The size of the “stanza window”, i.e., the number of codes preceding the reference unit can be selected by the researcher in the beginning of the analysis (on Figure 4b the moving stanza window size is

MSWS = 3). Once the algorithm starts, it takes each coded unit in the transcript one by one as a reference unit and tracks as many steps back as the size of the stanza window is (here MSWS = 3) in order to identify the connections. These steps are demonstrated on Figure 4b where each column represents a step during the movement of the stanza window. In the first column, the bold-framed code of “HG” represents the reference unit while the preceding codes of “EE, HG, HG” constitute the codes appearing in the stanza window. At this step, a connection between codes of “EE” and “HG” is observed. Then, the reference unit and the stanza window moves one step further (see the second column on Figure 3b) in order to make the same comparison between the reference line and the codes in the stanza window. This process ends once all coded units have been compared with preceding coded units. Then, further mathematical transformations and dimension reduction techniques (see e.g., Collier et al., 2016) are automatically conducted in order to extract network parameters that are identified by the relative frequencies of the overall co-occurrences of coded units. These network parameters (e.g., locations on nodes; strengths of connections between nodes; the centroids of the networks; Shaffer et al., 2016) are used for visualization and quantitative comparisons of different networks.

Epistemic networks have at least three relevant components for further analysis: nodes, connections between the nodes and centroids. *Nodes* in epistemic networks represent codes from the coded transcript (see e.g., “HG” on Figure 4a as code in the transcript and on Figure 4c as a “node” in a network). The *connections* between two codes indicate that those codes co-occurred over time in the discourse. More specifically, higher relative frequency of observed co-occurrences between two codes indicate stronger weighted connections between those codes compared to codes that do not co-occur that frequently throughout the discourse. The result (see e.g., Figure 5) is an overall network of all codes that occurred through the discourse, with connections between each other that become “thicker” the more often these codes occur together within the chosen stanza window size.

Furthermore, quantitative comparison between networks or groups of networks (e.g., networks of individuals vs networks of collaborating partners) is possible by using calculated *centroids* for every epistemic networks generated by ENA. A centroid is determined by the strength of connections between nodes in an epistemic network. The centroid value is calculated as the mean of connection weights (Shaffer et al., 2016) and as such, it represents all the weighted co-occurrences in a network on the discourse or on the group level (depending on whether the epistemic network represents e.g., a collaborating pair’s network or a network of a group of collaborating pairs. While this value provides information about the structural arrangement of any network, it allows the quantitative comparison of different networks (e.g., average network of collaborating partners vs average network of individuals) in the same space, based on the idea that the further two centroids fall from each other the more probably they represent different networks (Shaffer et al., 2016). Finally, qualitative comparison of epistemic networks is possible by interpreting visualizations. ENA offers different visualizations for that purpose, such as the visualization of networks (see e.g., Figure 5) or the visualization of subtracted networks. Subtraction of networks means that the values or strengths of nodes and connections of one network can be subtracted from another network. The resulting “subtracted network” represents the difference between the two networks and it can, thus, illustrate what makes e.g., collaborative networks of scientific reasoning different from individual networks of scientific reasoning.

ENA has been successfully used to identify networks of skills relevant for future practitioners in domains such as engineering (Arastoopour, Shaffer, Swiecki, Ruis & Chesler, 2016), journalism (Hatfield, 2015) or urban planning (Bagley & Shaffer, 2015). Moreover, by calculating networks that represent overall co-occurrences of reasoning processes in a discourse as well as in a group (average network of e.g., individual reasoners), it allows not only drawing valid conclusions regarding overall scientific reasoning, but also quantitative and qualitative comparisons between different groups of reasoners (e.g., collaborators vs individuals). Therefore, ENA seems to be an optimal methodological framework (Chi, 1997) to identify patterns of scientific reasoning and to compare groups (i.e., individuals and collaborating partners) of future practitioners regarding their scientific reasoning patterns.

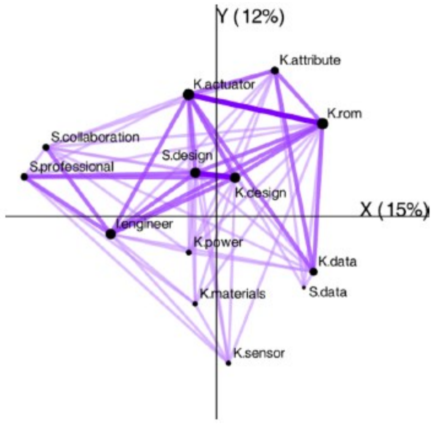

Figure 5. An exemplary epistemic network. Thicker nodes represent more frequent codes and thicker connections represent more frequent co-occurrences between those codes. Source: Shaffer et al. (2016, p. 3).

6. General Research Questions

The main aim of the present dissertation is to find out whether establishing a collaborative problem solving context can facilitate the engagement of future practitioners in scientific reasoning while they are solving authentic professional problems from their future practice. The introduction presented a novel framework on scientific reasoning (Fischer et al., 2014) and according to that framework it suggested that problem solvers can engage in scientific reasoning in two main ways: (1) by engaging in epistemic processes (process level of scientific reasoning) and (2) by applying scientific knowledge (the content level of scientific reasoning).

Empirical studies show that (future) practitioners show difficulties at both the process and the content levels of scientific reasoning (Hetmanek et al., 2015; Stark et al., 2009; van de Pol et al., 2011). Therefore, it is a reasonable question to what extent a minimal intervention approach may facilitate the engagement of future professionals in scientific reasoning processes and content use.

Research indicate that collaboration can foster the engagement in productive reasoning (Teasley, 1997; Chi, 2009) as well as scientific reasoning benefits from a collaborative compared to an individual inquiry context (e.g., Okada & Simon, 1997). Yet, the advantages of collaboration cannot always be manifest and difficulties between collaborating partners often occur (Clark & Brennan, 1991; Stasser & Titus, 1985). Besides, empirical research investigating the effect of collaboration on different processes and, at the same time, on the content aspect of scientific reasoning is scarce (e.g., Métrailler et al., 2008). Therefore, the present work investigates whether and in what respect collaboration, as a potential learning context to acquire scientific reasoning skills, might indeed improve scientific reasoning during professional problem solving. Accordingly, the following research questions are set to guide inquiry:

Research question 1: To what extent do collaborative reasoners engage in scientific reasoning processes and content use compared to individuals when solving authentic problems from their future professional practice?

To be able to find a reliable answer to the research question, it is important to take into account that the effect between collaborative and individual reasoning might be influenced by factors such as group heterogeneity regarding group members' prior script-like knowledge (Canham et al., 2012; Fischer et al., 2013). Therefore, dyadic heterogeneity on its members' problem solving scripts is considered for the second research question:

Research question 2: To what extent does group heterogeneity regarding its members' problem solving scripts influence the engagement in scientific reasoning processes and content use?

Moreover, scientific reasoning is a complex skill or set of skills (e.g., Osborne, 2010). Therefore, to validly capture the difference between scientific reasoning skills of collaborative vs individual reasoners, it is necessary to apply an appropriate measurement tool that affords to capture such complexity of reasoning skills. For this purpose, the present study chose the method of epistemic network analysis (Shaffer et al., 2009) that proved to be a reliable and valid method to identify reasoning patterns in different professional domains (e.g., Shaffer, Collier & Ruis, 2016). Accordingly, the third research question for the present work is set as:

Research question 3: What patterns of scientific reasoning can be identified that capture the distinctive features of collaborative vs individual scientific reasoning processes?

To answer these research questions, two empirical studies were conducted and these are going to be discussed in the following chapters. Study 1 (Chapter 8) is aiming to provide an analysis to empirically test Research question 1-2; while Study 2 (Chapter 9) is focusing on Research question 3.

7. Study 1: Pre-service teachers' scientific reasoning during pedagogical problem solving: better together?

7.1. Abstract

Many professions require that professionals solve practice problems by referring to scientific theories and evidence in a systematic, inquiry-oriented way. However, during their training, future practitioners often have problems in doing so. Collaborating with other future colleagues is discussed as a way that might help to establish more evidence-based problem-solving in practice. Although some theoretical and empirical evidence suggests that groups may outperform individuals in scientific reasoning, there is a lack of empirical evidence on comparing groups with individuals on reasoning scientifically when solving authentic problems from their practical domain (e.g, teachers reasoning about their pupil's performance). According to the Script Theory of Guidance, partners' engagement in collaborative reasoning is influenced by reasoners' initial scripts such as prior knowledge on how to solve problems. We therefore hypothesize that (1) collaborating partners might solve problems in a more scientific or evidence-based manner than individuals, and that (2) engaging in collaborative scientific reasoning is influenced by the heterogeneity of the collaborators' problem-solving scripts. In this study, 76 pre-service teachers solved an authentic problem case (reasoned about an underperforming student) from their future practice either individually or in dyads. Results show that dyads engaged more extensively in generating hypotheses, evaluating evidence, and they were also more likely to draw conclusions compared to individuals. On the other hand, dyads engaged less in generating solutions to the problem and referred to a lesser extent to scientific theories and evidence than individuals did. Furthermore, the more heterogeneous dyads were with respect to their problem-solving scripts, the less they engaged in generating solutions to the problem. These

results indicate that (1) scientific reasoning processes of pre-service teachers can benefit from reasoning about a problem with a partner. Also, (2) heterogeneity of groups regarding their problem solving approaches should be considered when planning collaborative reasoning sessions in higher education.

Keywords: evidence-based teaching; collaborative reasoning; problem-solving scripts; scientific reasoning

7.2. Research Problem

In many practical professions it is crucially important that practitioners make decisions in an evidence-based manner and engage in scientific reasoning. This means that when they solve problems, practitioners should (1) use relevant scientific findings from their domain (2) in a systematic inquiry-based manner similar to the way scientific knowledge is constructed (Fischer et al., 2014). For example, teachers solve pedagogical problems everyday such as how to spark their students' interest or deal with differences in performance (e.g., König, Blömeke, Paine, Schmidt, & Hsieh, 2011). To enhance the validity of their solutions, teachers need to apply psychological and didactical knowledge (see Baumert & Kunter, 2006) as well as to engage in a systematic, inquiry oriented way of reasoning by developing explanations about the problem and negotiating solutions for it (Fischer et al., 2014). Although developing general scientific reasoning skills is typically part of the curricula in higher education (e.g., pre-service teacher training), utilizing scientific reasoning for solving complex problems is frequently a problem for future practitioners (Gruber, Mandl, & Renkl, 2000).

We argue that one way to engage students during their academic training in scientific reasoning while solving authentic problems may be via encouraging them to collaborate while solving a problem instead of solving it alone (Baeten & Simons, 2014). Studies on collaborative learning and problem solving suggest that interactive discussions between learning partners can be more beneficial for learning outcomes and also for scientific reasoning than simply reasoning alone (Chi, 2009; Okada & Simon, 1997). Whether it is also the case for profession-related problem solving of future practitioners is the first main question of this article.

Nevertheless, problem solving groups can be very different from another regarding their members' knowledge; motivation; aims etc. that can all influence group performance. Indeed, future practitioners have diverse backgrounds and in addition, they might show very different approaches on solving complex problems and whether and how they would engage in scientific reasoning while solving problems. These approaches or "problem-solving scripts", according to the Script Theory of Guidance (Fischer, Kollar, Stegmann & Wecker, 2013), can have an effect on the groups' performance by guiding collaborating partners' engagement in scientific reasoning. More specifically, the more divergent strategies reasoning partners have, the more they may stimulate each other during solving a problem, or even the opposite: the more conflicts they may face. Whether or not such problem-solving scripts influence the engagement in scientific reasoning during collaborative problem solving is the second main question of this paper.

To summarize, the main aim of this study is to investigate whether scientific reasoning of pre-service teachers who solve an authentic problem from their future practice differs if they work (1) individually or as a group, and (2) if the group they are working in is homogeneous or heterogeneous with respect to the group members' problem-solving scripts.

7.3. Scientific Reasoning and Problem Solving

In many professions, problem solving requires the application of scientific theories and evidence from the domain in which the problem is grounded as well as systematic strategies or heuristics to approach the problem (Klopp, Stark, Kopp, & Fischer, 2013; Mullins, Rummel, & Spada, 2011; Rittle-Johnson, Siegler, & Alibali, 2001; Rummel & Spada, 2005; Weinstock, 2009). In some professions, this has been labeled "evidence-based practice" (e.g. (Satterfield et al., 2009). Fischer et al. (2014) argues that evidence-based practice can be seen as a specific instance of scientific reasoning: it can be understood as an engagement in distinct epistemic processes that are useful to construct knowledge about the causes of the problem and to derive possible solutions to it (see also Klahr & Dunbar, 1988; Zimmerman, 2000). In the context of this study, we differentiate between the *content* and the *process* level of scientific reasoning.

At the *content level*, problem solvers are expected to refer to relevant theories and empirical studies that have been conducted within their domain, as they are confronted with an authentic problem from their (future) professional practice (Gruber, Mandl & Renkl, 2000). For example, teachers and pre-service teachers should be able to apply psychological and educational theories of learning, motivation, instruction etc. and related research findings in order to solve a given problem (Hetmanek et al., 2015; Voss, Kunter, & Baumert, 2011), such as how to best support an underperforming student.

At the *process level*, scientific reasoning can be viewed as an inquiry process (Klahr & Dunbar, 1988) that is characterized by the engagement in epistemic processes reasoners follow while solving a problem (Fischer et al., 2014). Based on Fischer et al. (2014), we argue that when being confronted with problems from their actual practice, (future) professionals engage in a set of epistemic processes that are similar to those that scientists engage in when solving a research problem. They need to identify the problem itself (Problem Identification); ask questions or make statements that guide their further exploration on the problem (Questioning); set up candidate explanations for the problem (Hypothesis generation); take into account or generate further information necessary to understand or solve the problem

(Evidence Generation); evaluate the information in the context of their hypotheses (Evidence Evaluation); plan interventions / solutions (Constructing artefacts); engage in discussions with others to re-evaluate their thoughts (Communicating and scrutinizing) and sum up their process to arrive at well-warranted conclusions on how to explain and/or solve the problem (Drawing conclusions).

Nevertheless, as prior research has shown, (future) professionals seem to have difficulties in both the content and process aspects of scientific reasoning. For example, in-service and pre-service teachers often seem to neglect alternative explanations for given problems, frequently use scientific theories and models only in an insufficient manner (Sampson & Blanchard, 2012; Stark, Puhl, & Krause, 2009) and have difficulties of applying their knowledge to solve educational problems (McNeill & Knight, 2013).

Similarly, when teachers assess learning processes, for example, by watching videos of teacher-student interactions in the class, they seem to show relatively lower engagement in generating evidence-based hypotheses on their own compared to their performance after a methods course (Yeh & Santagata, 2015). This suggests that teachers could potentially reason better than they do, if they are supported in the right way – and this refers to both the content and the process level. With respect to the content level, Lockhorst, Wubbels and van Oers (2010) found that in their study teachers indicated a lack of scientific knowledge to assess pupil's learning processes. With respect to the process level, van de Pol, Volman, & Beishuizen, (2011) showed that teachers often fail to collect enough evidence to properly scaffold the work of their pupils.

Taken together, in many domains, (future) professionals show difficulties with regard to both components of SR, i.e. engaging in epistemic processes and using scientific content knowledge (Zimmerman, 2000), when it comes to solving authentic problems from their (future) professional practice. Thus, searching for ways to help (future) professionals in this process in order to prepare them for a more evidence-based future practice is urgently needed.

7.4. Collaborative Problem Solving and Scientific Reasoning

One potential way to help future professionals to engage in scientific reasoning is to ask them to collaborate while solving complex problems. Psychology and learning research suggest that collaborative problem-solving can have benefits over individual problem-solving (Barron, 2000; Chi, 2009; Chi & Wylie, 2014; Kirschner, Paas, Kirschner & Janssen, 2011; Mullins, Rummel & Spada, 2011). Reasoning as a group has the advantage of articulating and building on knowledge each collaborating partner brings to explain and solve a problem (Asterhan & Schwarz, 2009; Chi & Wylie, 2014; Elbers & Streefland, 2000; Hausmann, Chi, & Roy, 2004; Kirschner, Paas, Kirschner, & Janssen, 2011; Teasley, 1995). Teasley (1995), for example, has demonstrated that in a problem solving context, children showed more engagement in productive reasoning when they reasoned together. Namely, dyads' engaged significantly more in hypothesizing than individuals and produced proportionally more interpretive talk than, while individuals did. In another study (Okada & Simon, 1997), university undergraduates participating as dyads or individuals showed a similar outcome on a scientific reasoning task. In this study, dyads engaged more in hypothesis generation and evidence evaluation (planned experiments to test the initial hypothesis) than individuals did. As the authors note, one reason for such heightened engagement in explanatory behavior may be the need to communicate in a more explicit manner (Okada & Simon, 1997). More recent studies with psychology undergraduates (Métrailler, Reijnen, Kneser & Opwis, 2008), found that although dyads may be proportionally less concerned about hypotheses than individuals, but still engage more in explanatory behavior (e.g., disagreeing with each other). Such dialectic turns occurring in a collaborative scenario can also be part of a coordination attempt for meaning-making and knowledge co-construction (Fischer & Mandl, 2005; Hausmann, Chi & Roy, 2004) which is a relevant aspect of scientific reasoning (Barron, 2000; Fischer & Mandl, 2005; Roschelle, 1992; Roschelle & Teasley, 1995). Such collaborative knowledge co-construction can eventually lead to deeper processing of knowledge and better problem solving performance (Chi & Wylie, 2014; Hausmann, Chi & Roy, 2004; Kirschner et al., 2011).

Collaboration among professionals such as teachers and pre-service teachers, when they solve professional (e.g., pedagogical) problems together rather than individually, might offer great benefits too (Baeten & Simons, 2014; Bullough et al., 2002). Collaborating teachers have potentially more opportunity to engage in professional discourse: to exchange different perspectives and methodological knowledge on teaching and learning which is expected to facilitate their growth on becoming reflective and evidence-based practitioners (Baeten & Simons, 2014; Birrell & Bullough, 2005; Gardiner & Robinson, 2009; Kamens, 2007; Nevin, Thousand, & Villa, 2009; Nokes, Bullough Jr., Egan, Birrell, & Merrell Hansen, 2008). In a field study, for example, Nokes et al. (2008) investigated whether preservice teachers benefit from paired teaching. Pre-service teachers in this study had to teach in a secondary school for a 15 weeks period. The qualitative interviews with the teachers and their mentors indicated that collaboration gave them the opportunity to engage in dialogue, share their experiences and become more reflective about their teaching practices.

Yet, collaboration does not always work and does not always lead to better reasoning processes and outcomes compared to individual reasoning. Groups may suffer from process losses due to coordinational difficulties, such as disagreements on how to move on. Such coordinational challenges can result in poor performance, e.g., in problem-solving tasks (Diehl & Stroebe, 1987; Strijbos, Martens, Jochems, & Broers, 2004; Weinberger, Stegmann,

& Fischer, 2010). Indeed, unclarified actions can slow down or hinder the development of mutual understanding and constructive communication processes (Horton & Gerrig, 2005; Roschelle & Teasley, 1995; Rummel & Spada, 2005). Yet, excessive reliance on shared knowledge without discussing unique ideas (e.g., in order to avoid disagreements) has been described as another problem that can hinder groups to reach their potential by building on the diverse knowledge of its members (Stasser & Titus, 1985).

Difficulties with coordination or lack of shared knowledge between collaborators are among those problems that have been also reported in the context of professional problem solving (e.g., Rummel & Spada, 2005). In team-teaching, for example, planning the lesson with a partner requires coordination between the teachers (Jang, 2008; Nokes et al., 2008). This might, especially in case of disagreements between reasoning partners, lead to difficulties (Bullough et al., 2002): collaboration might become time consuming (Jang, 2008) and contribute to a higher perceived workload (Nokes et al., 2008). In their review, Baeten and Simons (2014) lists the lack of compatibility (e.g., different conceptions of teaching), social comparison (that one would outperform the other), the difficulty of providing constructive feedback and the perceived increase in workload among the barriers of team-teaching. In another review, Shin, Lee, & McKenna, (2016) note that all of their reviewed studies reported some challenges for preservice teachers in collaborative settings. They mention it as problematic that many preservice teachers were unprepared for collaborative practices; it was unclear for them what they were supposed to do and often they lacked the skills of instructional planning which served as a barrier for fruitful collaboration.

All in all, although the cost of coordinating with a partner may require increased efforts, exchanging different views is a clear benefit of it, resulting in engagement in dialogues and professional growth (Birrel & Bullough, 2005; Nokes et al., 2008). Yet, whether groups really can benefit from the fact that they are working together may heavily depend on the way groups are formed. This issue is being explained in the following section.

7.5. How to Form Groups: about the Heterogeneity of Collaborators' Problem-Solving Scripts

Within collaborative learning research, there have been long debates about the question whether group members should be similar (i.e., homogeneous) or dissimilar (i.e., heterogeneous) to each other (e.g., Bowers, Pharmer, & Salas, 2000). On the one hand, in heterogeneous groups, reasoners may stimulate (Paulus, 2000) each other by bringing new perspectives to a problem solving task (Miura & Hida, 2004; Stroebe & Diehl, 1994). A meta-analysis by Bowers, Pharmer and Salas (2000) concludes that homogeneous teams tend to perform better on rather routine-like low-difficulty tasks that are clearly defined and require simple responses (e.g., solving a puzzle) while heterogeneous teams perform better on more complex problem solving tasks that require more cognitive resources (e.g., business games). More recent empirical studies (Canham et al., 2012; Wiley et al., 2013) also support the notion that group heterogeneity is beneficial for innovative problem solving. On the other hand, heterogeneous groups may need more time and effort to reach a common understanding of the problem and coordinating their strategies to solve it (Bullough et al., 2002; Rummel & Spada, 2005). For example, in teacher collaboration diverse approaches for classroom management has been reported rather problematic (Shin et al., 2016).

Overall, available empirical studies addressing group heterogeneity are not conclusive regarding in what aspects groups should be heterogeneous. They lack a joint framework on how heterogeneity can be operationalized and measured across studies (Van Knippenberg & Schippers, 2007).

Indeed, groups can be homo- or heterogeneous in a number of different ways. While previous studies focused on prior knowledge (Wiley & Jolly, 2003; Wu, Liao, & Dai, 2015), demographic background (Curseu & Pluut, 2013; Wegge et al., 2008), or attitudes (Curseu & Pluut, 2013), the present article is particularly interested in group heterogeneity of the collaborators' "problem-solving scripts", which define as their knowledge and expectations on how to reason about problems from professional practice (Fischer et al., 2013). For example, clinical reasoning scripts, e.g., "illness scripts" (Charlin, Boshuizen, Custers, & Feltovich, 2007) or "ward round scripts" (Beltermann, Wessels, Kollar, & Fischer, subm.) can guide medical professionals diagnostic behavior. A physician meeting a pale patient can identify that there may be a problem, and after engaging in a systematic evidence generation procedure (e.g., listening to the symptoms, conducting further investigations), she can revise the plausible hypotheses (diagnoses) that can answer the original problem (Charlin et al., 2007). Similarly, in the domain of teaching, when faced with underperforming students, one teacher may tackle such problems by systematically analyzing the problem and deliberately searching for educational or psychological theories to make inferences regarding promising actions. As part of the process, she may generate hypotheses about the reasons for the problem (e.g., heterogeneity in abilities), evaluate evidence (e.g., the problem is specific to the given situation, therefore ability might not play the decisive role), and generate solutions (e.g., the intervention should target the specific situation or the students' perception of such situations). Yet, another teacher may solve such problems by mentally going through past experiences and quickly jump to conclusions without evaluating the available information further. For example, she might think of a friend in school who behaved in a similar problematic manner than her present problem case and draw conclusions analogous to that friends' case. We assume that such "reasoning scripts" (see also Schank, 1999) direct the way reasoners understand and act in a given problem-solving situation (Weinstock, 2009).

Although the Script Theory of Guidance (Fischer, Kollar, Stegmann & Wecker, 2013) suggests that reasoners' individual (problem-solving) scripts may have an impact on their collaborative reasoning (Kollar, Fischer, & Slotta, 2007); the effects of the level of heterogeneity of problem-solving scripts within groups on reasoning processes while solving authentic problems has hardly been empirically addressed (Vogel et al., 2016). Scarce attempts (Pavitt & Johnson, 2001) have taken into account individuals' scripts on group problem solving processes; yet, those studies have not investigated whether the level of group heterogeneity influences certain problem solving processes.

7.6. Research Questions

In our empirical study, we were interested in whether and how pre-service teachers engage in scientific reasoning while solving authentic problems from their future professional practice. More specifically, we were interested in the question whether (a) groups (dyads) would refer more or less to scientific theories and evidence and perform epistemic processes of (scientific) reasoning more extensively than individuals and whether (b) homo-/heterogeneity of dyads' problem-solving scripts might influence to what extent dyads refer to scientific content and perform certain epistemic processes of (scientific) reasoning. Thus, we set up the following two research questions:

RQ1: Do dyads of preservice teachers differ from individual preservice teachers in the extent to which they (a) refer to scientific theories and evidence and (b) in their engagement in different epistemic processes of scientific reasoning?

RQ2: Within dyads, does heterogeneity regarding group members' problem-solving scripts affect the extent to which dyads of preservice teachers (a) refer to scientific theories and evidence and (b) engage in different epistemic processes of scientific reasoning?

7.7. Method

7.7.1. Participants and Design

76 teacher education students (59 female, $M_{\text{Age}} = 21.22$, $SD = 3.98$) from a German university (on their first to fifth semesters) participated in the study. They received course credit for their participation. To answer research question 1, each participant was randomly assigned, in a between-subject-manner, to either an individual (16 students, 13 female, $M_{\text{Age}} = 22.31$; $SD = 6.73$) or a dyadic (60 students, i.e. 30 dyads, 46 females, $M_{\text{Age}} = 20.93$; $SD = 2.85$) condition. To answer research question 2, an index was calculated to measure each dyad's heterogeneity level based on a test in which students had to individually describe how they approach authentic problems from educational practice (see below). The resulting score on the heterogeneity index was then used as a predictor for the use of scientific theories and evidence as well as for the dyads' engagement in the different epistemic processes of scientific reasoning.

7.7.2. Procedure

The procedure of the study consisted of four steps. Regardless of the condition (dyadic vs individual), in the first three steps every student participated individually. In the last (problem solving) step, students participated either as dyads or individually depending on the condition they were assigned to. First, students filled in a computer-based questionnaire on demographic variables. After that, they were given a computer-based card sorting task to measure their problem-solving scripts (see below). Then, they were given five minutes to read five printed out presentation slides that included scientific content information originating from their introductory psychology class and short descriptions of theories and concepts (e.g., on strategic use of short-term memory and a classification of learning strategies). After that, participants were presented an authentic problem from their future professional practice, which was: “You are a teacher in a school. One of your students receives low grades in comparison to others. The student looks motivated and it seems she understands the content. You know from the parents that she learns diligently at home. You as a teacher, please find possible reasons and a solution to the problem”. For dealing with the problem, participants had 10 minutes. During these ten minutes, students in the individual condition were asked to think aloud while solving the problem; dyads were asked to orally discuss the problem. All think-aloud and collaborative discussions were audio-recorded and transcribed for analysis (see below). Data from this problem-solving process were used to measure students’ use of scientific theories and evidence as well as their engagement in the epistemic processes of scientific reasoning (see below). Finally, students were thanked for their participation and debriefed. The whole data collection took about one hour for each individual or dyad.

7.7.3. Independent Variables

Reasoning setting. Reasoning setting was varied by randomly assigning participants to the problem-solving phase either as individuals or dyads. All dyads were randomly established. Thus heterogeneity level was identified via a post-hoc assessment.

Heterogeneity of reasoning scripts within dyads (dyadic heterogeneity). We defined heterogeneity of dyadic composition by comparing dyadic members’ problem-solving scripts that were measured initially during the card sorting task that preceded the problem-solving phase. During the card sorting phase, students participated individually. First, they were presented the practice-related problem case described above. Then they were asked to use a set of prefabricated cards available on a MS PowerPoint slide to indicate what (epistemic) processes they would perform while solving the presented problem. The eight epistemic processes from Fischer et al. (2014) and five additionally selected “distractor” processes (e.g., “Giving feedback”, “Improvising”) were written on the cards that were presented to the participants. Besides that, five blank cards were provided to give participants the opportunity to note down further processes if they wanted to. From the resulting process sequences, participants’ problem-solving scripts were coded according in the following way: We summed those epistemic process cards representing processes from the Fischer et al. (2014) model that were selected by both dyadic members. This number represented their shared knowledge component index (SKCI) on scientific reasoning. Then, we calculated a disagreement on position index (DPI) between dyadic members by calculating how many out of the epistemic processes they agreed on (SKCI) would need to be switched in position so as both members’ selection shows the same sequence of the shared epistemic processes on scientific reasoning. We also calculated a pooled knowledge component index (PKCI) on scientific inquiry by summing the number of epistemic processes of the Fischer et al. (2014) model that at least one dyadic member had selected. Finally, a homogeneity index was calculated as (SKCI - DPI)/PKCI to account for agreements and, at the same time controlling for disagreements between the two members of a group. Larger values of this index indicated more dyadic homogeneity, while lower values indicated more dyadic heterogeneity.

7.7.4. Dependent Variables

Dependent variables were collected during the problem-solving phase that had students (dyads or individuals) solve the aforementioned authentic problem from their future professional practice (see above). To make SR processes accessible for research, we asked students in the individual condition to think aloud and students in the dyadic conditions to verbally discuss how they would solve the problem.

All verbal data were transcribed. Before coding, we segmented the transcribed data into syntactical proposition-sized units (Chi, 1997). 10% of the data was independently segmented by two researchers after a training on segmentation. Reliability was calculated as the proportion of agreement according to Strijbos, Martens, Prins, & Jochems (2006). Thus, to more precisely detect reliability, agreements were computed from both segmenters' perspective: In 85.09% out of the total number of segment boundary indicated by Segmenter 1, Segmenter 2 also agreed that there is a segment boundary. At the same time, in 79.73% out of the total number of segments indicated by Segmenter 2, Segmenter 1 also indicated a segment boundary. These values indicated the reliability of the segmentation scheme. In a next step, one of the segmenters segmented the remaining data. We developed coding schemes to assess students' engagement in (a) use of scientific theories and evidence and (b) epistemic processes of scientific reasoning.

Use of scientific theories and evidence. A second coding scheme was developed to capture for each segment whether or not participants used scientific theories and/or evidence or not. Segments were coded as "use of scientific theories and evidence"" if the speaker referred to scientific theories, concepts or methods. Specifically, we used the following five categories: "Learning strategy" was applied when participants mentioned ways of processing the learning material ("Elaborated strategy"; "Recollection difficulties") related topics. "Anxiety" was applied when participants referred to test anxiety or emotional pressure. "Motivation" was used when participants talked about motivation. As the last scientific content code we used the category "Other" such as "Self-fulfilling prophecy"; "Mobbing"; "Mind-map". If the above codes did not apply, the segment was coded as non-scientific content related. For each transcript (individual or dyadic), we merged all content categories to calculate overall engagement in use of scientific theories and evidence, then we divided this value with the total number of propositions in order to calculate our variable for the analyses. 5% of the segments were coded by two independent coders with an agreement of Cohen's κ=.82.

Epistemic processes. All segments were coded by aid of a coding scheme (see Table 1) developed to capture the eight epistemic processes of scientific reasoning proposed by Fischer et al. (2014) or non-epistemic propositions. Since it proved impossible to reliably differentiate between the two activities "evidence generation" and "evidence evaluation", we merged these two categories into one: evidence evaluation. After the coding scheme had been developed and trained, two independent coders coded 10% of the data for the identification of epistemic processes. Inter-rater reliability was sufficient (Cohen's κ = .68). After coding each segment, the numbers of segments that fell in the same category were summed for each epistemic process: problem identification (PI), questioning (Q), hypothesis generation (HG), generating solutions (GS), evidence evaluation (EE), drawing conclusions (DC), communicating and scrutinizing (CS), and non-epistemic proposition (NE). Each resulting sum scores were divided by total talk to include in the statistical analyses. Finally, as the data screening (see below) revealed that three epistemic processes (problem identification, questioning and drawing conclusions had relatively large non-coded ratio, we dummy-coded these variables by assigning 0 if there was no proposition coded under the given epistemic process and assigning 1 if there was at least one proposition coded under that epistemic process.

Table 1
Coding Scheme to Capture Epistemic processes of Scientific Reasoning.

Code for Epistemic process	Brief Description of Code	Example of Code
Problem identification (PI)	An initial attempt to build an understanding of the problem.	"So it is about a student, // who has low grades"
Questioning (Q)	A question or statement orienting further inquiry.	"Ok, so what is the reason for that?"
Hypothesis generation (HG)	Any explanation of the problem case.	"So if the reason is her learning method"
Evidence generation (EG) (later merged with EE)	Referring to case information;	"She studies diligently at home"
	to scientific evidence;	"There are different learning strategies…"
	to anecdotal evidence;	"I know someone who has exam nerves"
	to lack of information;	"We also do not know her age."
	or planning further information collection.	[to find out] "how much time she needs to do her homework. "
Evidence evaluation (EE) (later merged with EG)	Evaluation of evidence (to support / falsify HG or CA.	"and then you can even exclude the problem of exam nerves"
Generating solutions (GS)	Planning an intervention, how to solve the problem.	"You should discourage her from using surface strategies"
Communicating & scrutinizing (CS)	Planning to engage others in the inquiry process.	"You can also talk to the parents"
Drawing conclusions (DC)	Concluding the outcomes of the earlier steps of inquiry.	"For me these would be the most important points to understand at all what her problem is."
Non-epistemic (NE)	Propositions that cannot be coded under the other codes.	"Ok, have you read it through?"

7.7.5. Statistical Analyses

To answer the question whether dyads differ from individuals on their engagement in their use scientific theories and evidence and in their engagement in epistemic processes (RQ1) ANOVAs and a MANOVA were conducted. For solution generation a Welch-test was conducted. For the dummy-coded variables on epistemic processes of problem identification, questioning and drawing conclusion, chi-square tests were applied.

To answer the question if dyadic heterogeneity has an effect on preservice teachers' use of scientific theories and evidence and their engagement in epistemic processes (RQ2), we analyzed only data that came from dyads and conducted linear regressions with "dyadic heterogeneity" as predictor and "use of scientific theories and evidence" as well as "engagement in SR activities" as criterion variables. For problem identification, questioning and drawing conclusion logistic regressions were conducted.

For all analyses, the unit of analysis was the transcript, no matter whether it came from an individual or a dyad. I.e., transcripts from individual think-aloud transcripts were compared with transcripts from dyadic discussions.

For all analyses, the alpha level was set to $p < .05$.

7.8. Results

7.8.1. Data Screening and Diagnostic Statistics

For the first analysis (effect of reasoning setting on scientific content use), z-scores for skewness and kurtosis were calculated (Gardiner & Robinson, 2009, pp. 138–139) to investigate the normality distribution within the groups of comparison. Dyads showed a positively skewed ($z = 3.87$) as well as a leptokurtic ($z = 4.80$) deviation from normality. Data screening indicated one outlier case ($z = 3.50$), that also accounted for the skewness in the data, and was thus removed from further analysis to lower a potential Type II error rate

(Osborne & Overbay, 2004). However, Levene's test indicated unequal variances between individuals and dyads ($p < .05$), therefore a Welch's F-test was conducted due to its robustness (i.e., its affordance to keep Type I error at a satisfactory level) for the violation of homogeneity of variances in case of unequal sample sizes (Field, 2009, p. 379-380; Grissom, 2000).

For the second analysis (effect of reasoning setting on engagement in epistemic processes), Mahalanobis distances did not indicate multivariate outliers ($p \geq .03$). Bivariate scatterplots indicated linearity among most of the epistemic processes. The frequency distributions of three epistemic processes, however, suggested very high frequencies of missing values for the variables *drawing conclusions* (67.39%), *questioning* (58.70%) and *problem identification* (39.13%). Consequently, we decided to exclude these variables from further parametric analyses, rather including them as dummy-coded variables in alternative statistical tests (see above). Data screening showed no outliers for the remaining epistemic processes, neither for individuals nor for dyads ($z \leq 2.85$). Z-scores for skewness and kurtosis did not indicate a violation of normality for epistemic processes in case of dyads ($z \leq 2.07$). Bartlett's sphericity test showed the expected covariance among the remaining epistemic processes, Approx. $\chi^2(10) = 88.25$, $p < .001$. Testing multicollinearity showed only one very high ($|r| > .8$) negative bivariate correlation between solution generation and evidence evaluation in case of individuals, $r(15) = -.87$. However, since these two variables were not redundant for dyads, and we did not have a theoretical reason to exclude either of them from the analysis, rather to include both, we decided to run one analysis with both variables instead of two separate analyses with each. Box's test suggested equal covariance matrices ($p = .05$). Finally, Levene's test suggested equal variances in case of epistemic processes except for solution generation, where individuals showed significantly higher variance than dyads ($p <$

.05). Considering the assumption checks, we decided to conduct MANOVA with a robust statistic, such as Pillai's trace (Meyers, Gamst, & Guarino, 2006, p. 378). Also, we calculated Welch's F-test as a follow-up test for solution generation (Field, 2009).

For the third analysis (effect of dyadic heterogeneity on the use of scientific theories and evidence), z-scores for skewness and kurtosis indicated no deviation from normality for the distribution of dyadic heterogeneity ($|z| < .07$). As for scientific content use, the abovementioned outlier case was removed from the analysis. No further violation of the assumptions were found.

For the fourth analysis (effect of dyadic heterogeneity on the engagement in epistemic processes), we excluded the variable *non-epistemic propositions* from multivariate analysis due to its curvilinear relationship with dyadic heterogeneity. As for the univariate effects, the casewise diagnostics did not reveal any influential cases ($\hat{y} \leq .16$) or residual outliers out of a standardized residual boundary of $|z|=3$.

7.8.2. Results

RQ1: Do dyads of preservice teachers differ from individual preservice teachers in the extent to which they (a) refer to scientific theories and evidence and (b) in their engagement in different epistemic processes of scientific reasoning?

To answer the question whether dyads differ from individuals with respect to the extent to which they refer to scientific theories and evidence, a Welch's F-test was conducted with scientific content use as the dependent variable and reasoning setting as the independent variable. Reasoning setting showed a significant effect on the *use of scientific theories and evidence*, Welch's $F(1, 23.51) = 2.64$, $p < .05$, partial $\eta^2 = .17$ (see Figure 6). Individuals referred proportionally more (M=.44, SD=.20) to scientific content than dyads (M=.29, SD=.14) did.

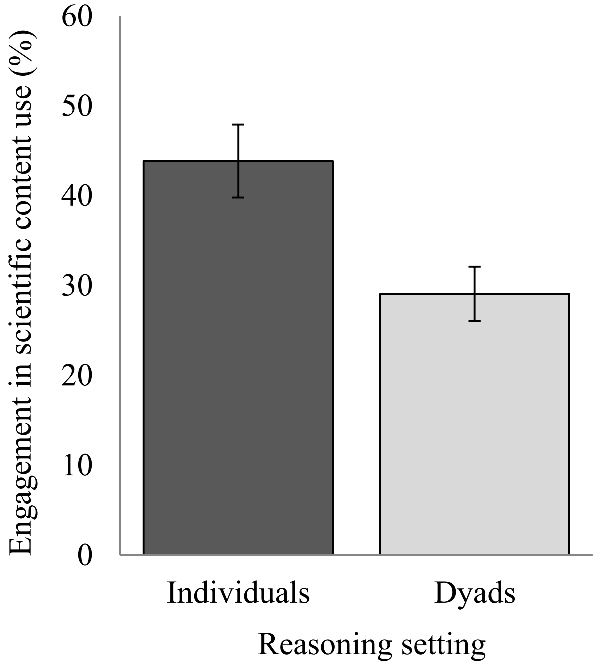

Figure 6. Use of scientific theories and evidence by individuals and dyads. Engagement was measured by the proportion of propositions to overall talk (here expressed as %). Error bars denote standard error around the mean.

Furthermore, to answer the question whether dyads differ from individuals on their engagement in epistemic processes, a MANOVA was conducted with the seven epistemic processes and non-epistemic propositions as dependent variables while reasoning setting (individual vs. dyadic) was included as the independent variable. Reasoning setting had a significant strong multivariate effect on the engagement in epistemic processes, Pillai's trace = .40, $F(5,40) = 5.26$, $p < .001$, partial $\eta^2 = .40$.

Follow-up ANOVA-s and a Welch's t-test for solution generation revealed significant effects of reasoning setting on the engagement in *hypothesis generation*, $F(1,44) = 6.06$, $p < .05$, partial $\eta^2 = .12$; in *evidence evaluation,* $F(1,44) = 4.28$, $p < .05$, *partial* $\eta^2 = .09$; in *solution generation,* Welch's $F(1, 19.79) = 6.56$, $p < .05$, 95%, partial $\eta^2 = .17$; as well as on *non-epistemic propositions*, $F(1,44) = 10.48$, $p < .01$, partial $\eta^2 = .19$. Figure 7 demonstrates the results. Dyads engaged more in *hypotheses generation* ($M = .24$, $SD = .09$) than individuals did ($M = .17$, $SD = .11$); they also engaged more in *evidence evaluation* ($M = .33$, $SD = .11$) than individuals ($M = .26$, $SD = .13$); and they made more *non-epistemic propositions* ($M = .06$, $SD = .04$) than individuals ($M = .02$, $SD = .03$). Yet, individuals engaged more in *solution generation* ($M = .45$, $SD = .24$) than dyads ($M = .29$, $SD = .13$).

Finally, a chi-square test indicated a significant relationship between reasoning setting and engagement in *drawing conclusions*, $\chi^2(1) = 4.51$, $p < .05$. The odds of engaging in drawing conclusions were 5.43 times higher for dyads than it was for individuals. No further significant effect was found.

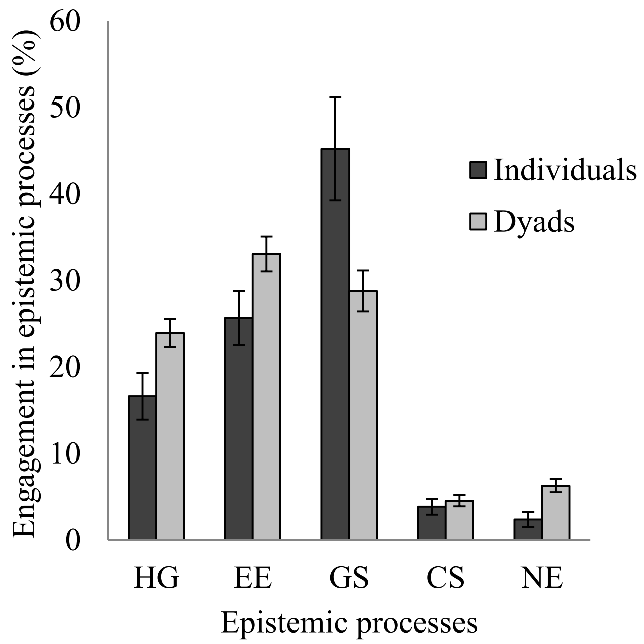

Figure 7. Engagement in epistemic processes (for epistemic processes corresponding to abbreviations see Table 1) by individuals and dyads. Engagement was measured by the

proportion of propositions to overall talk (here expressed as %). Error bars denote standard error around the mean.

RQ2: Within dyads, does heterogeneity regarding group members' problem-solving scripts affect the extent to which dyads of preservice teachers (a) refer to scientific theories and evidence and (b) engage in different epistemic processes of scientific reasoning?

To answer the question if dyadic heterogeneity has an effect on preservice teachers' engagement in scientific content use, we conducted a linear regression analysis with dyadic heterogeneity as the predictor and scientific content use as the criterion variable. Dyadic heterogeneity did not predict significantly the *use of scientific theories and evidence*, B = -.17, 95% CI_B = [-.65; .31], β = -.14, p = .47, R^2 = .02, *adj* R^2 = -.02.

To answer the question whether dyadic heterogeneity had an effect on the engagement in epistemic processes, separate linear regression analyses were conducted that included the dyadic heterogeneity as a predictor and solution generation, hypothesis generation, evidence evaluation, communicating and scrutinizing separately as criterion variables. The regression models revealed that dyadic heterogeneity significantly and positively predicted *solution generation*, B = .49, 95% CI_B = [.11; .86], β = .45, $p < .05$, R^2 = .20, *adj* R^2 = .17. The more heterogeneous dyads were the less they engaged in generating solutions. Moreover, dyadic heterogeneity negatively, yet, non-significantly predicted the following epistemic processes: *hypothesis generation*, B = -.17, 95% CI_B = [-.46; .11], β = -.23, p = .22, R^2 = .05, *adj* R^2 = .02; *evidence evaluation*, B = -.25, 95% CI_B = [-.60; .09], β = -.27, p = .14, R^2 = .08, *adj* R^2 = .04; and *communicating and scrutinizing*, B = -.04, 95% CI_B = [-.15; .07], β = -.14, p = .46, R^2 = .02, *adj* R^2 = -.02. To see whether dyadic heterogeneity has an effect on the overall proportional engagement in hypothesis generation and evidence evaluation, we summed the epistemic processes of generating hypothesis and evaluating evidence and regressed this summed value on dyadic heterogeneity. Dyadic heterogeneity significantly and negatively predicted the summed engagement in generating hypotheses and evaluating evidence, B = -.43, 95% CI_B = [-.79; -.06], β = -.42, p < .05, R^2 = .17, *adj* R^2 = .14. Finally, logistic regressions suggested no effect of dyadic heterogeneity on *problem identification*, *questioning* or on *drawing conclusion*.

7.9. Discussion

7.9.1. General discussion

The first aim of our study was to find an empirical answer to the question whether asking future practitioners (pre-service teachers) to collaborate is a useful way to engage them in scientific reasoning when they solve problems from their future profession. Moreover, we were interested in if accounting for group composition during that collaboration, i.e., regarding collaborating partners' problem solving scripts, can bring an additional advantage that might positively affect collaborative scientific reasoning. To be able to answer these questions, we differentiated between *content* and *process* aspects of scientific reasoning. So far there is a lack of empirical evidence to draw any conclusion whether pairing teachers could be advantageous for engaging them in scientific reasoning while solving pedagogical problems. Besides the theoretical predictions (e.g., Chi, 2009), empirical research on collaborative learning reported that groups might do better on problem solving than individual reasoners (e.g., Kirschner et al., 2011), while unshared ideas (Stasser & Titus, 1985) or coordination difficulties (e.g., Weinberger, Stegmann & Fischer, 2010) may limit their potentials. In accordance, qualitative research with preservice teachers (Baeten & Simons, 2014; Nokes et al., 2008) and research on scientific reasoning (e.g., Okada & Simon, 1997) indicate that collaboration is likely to result in more reflective discourse and higher engagement in epistemic processes such as hypothesis generation or evidence evaluation compared to individual reasoning. Finally, empirical research on problem solving suggest that group composition might positively affect collaborative problem solving (Bowers et al., 2000; Wiley et al., 2013) while heterogeneity of dyads may also result in an increased need to coordinate with each other (Bullough et al., 2002; Rummel & Spada, 2005).

RQ1: Do dyads of preservice teachers differ from individual preservice teachers in the extent to which they (a) refer to scientific theories and evidence and (b) in their engagement in different epistemic processes of scientific reasoning?

The first main finding of our study on the *content* aspect of scientific reasoning suggests that although pairs of teacher students might bring diverse knowledge on scientific theories and evidence to solve the problem, they do not seem to talk more about that. In the opposite, in our study dyads engaged proportionally less in discussing scientific theories and evidence than individuals did. One possible explanation of this finding is that dyads might be hesitant to discuss unshared content knowledge with each other (Stasser & Titus, 1985), and instead, they discuss content that is familiar for both of them. In a collaborative context, there are different cognitive and motivational aspects that might account for the engagement in discussing unshared knowledge (e.g., Paulus, 2000). Studies report that teacher students are typically not prepared for collaborating with each other (Baeten & Simons, 2014; Shin, Lee, & McKenna, 2016) . Therefore, it is plausible that in a collaborative situation the commonly known content can be a more “convenient” choice for them in the sense that it can help them to establish a joint understanding of the problem (Roschelle & Teasley, 1995) and hence, serve as a good basis for further collaboration. Such epistemic “closure”, however, has its own costs such as not going beyond shared knowledge, even if this knowledge does not stem from scientific sources.

In opposition to the results above, our second main finding on the *process* aspect of scientific reasoning was that dyads of pre-service teachers engaged more in epistemic processes of hypothesis generation (explaining the problem) and evidence evaluation than individuals did. Moreover, dyads were also more likely to draw final conclusions and to sum up their reasoning processes. These findings are not only consistent with prior research on scientific reasoning (e.g. Okada and Simon, 1997), but also with the theoretical assumptions of Fischer et al. (2014). Namely, that epistemic processes typically captured in scientific inquiry tasks (Klahr & Dunbar, 1988; Okada & Simon, 1997) can be generalized to a professional problem solving context. In other words, the scientific quality of professional problem solving seems to benefit from collaboration. Collaborating teachers can be more reflective and evaluative compared to individual practitioners, such as earlier qualitative findings in team-teaching also suggest (Baeten & Simons, 2014; Nokes et al., 2008).

This might come, however, at the cost of engaging less in generating solutions to the problem. Comparing this finding with the previous one on reasoning processes, it seems that there is a trade-off between explanatory and solution processes when comparing collaborative and individual problem solving. It appears that the lower engagement in solution generation in case of dyads can be explained by their enhanced engagement in hypothesis generation and evidence evaluation. One explanation is that the epistemic processes of giving explanations, referring to evidence and sum up the discussion reflect to the epistemic need to coordinate between reasoning partners to develop a better (shared) understanding of the problem (Rummel & Spada, 2005). As Okada and Simon (1997) argue, that in a collaborative situation reasoners must be often more explicit to communicate their point to make their partner understand and/or accept their ideas. Furthermore, it is also possible that although an enhanced engagement in explanatory processes may serve coordination purposes between reasoning partners, a decreased involvement in solution generation might also reflect different epistemic aims or criteria (e.g., enhanced need for validity) that may surface while discussing with others as compared to thinking about the problem individually. Further studies could potentially extend on our work by investigating whether (pre-service) teachers might indeed have different epistemic criteria when solving problems than individual teachers and to what extent that might influence their problem solving processes and outcomes.

RQ2: Within dyads, does heterogeneity regarding group members' problem-solving scripts affect the extent to which dyads of preservice teachers (a) refer to scientific theories and evidence and (b) engage in different epistemic processes of scientific reasoning?

Based on earlier findings (Bowers et al., 2000), we assumed that collaborative problem solving might benefit from dyadic members' heterogeneous problem solving scripts, while script heterogeneity might also lead to increased coordination demands for dyads.

As for the *content* aspect of scientific reasoning, dyadic heterogeneity on collaborating partners' problem solving scripts did not seem to affect the use of scientific theories and evidence. Considering the results between individuals and dyads it is possible that dyadic heterogeneity did not show an impact of using scientific content for similar reasons. Namely, although members of the groups may bring unique knowledge to a group discussion, groups might tend not to share that knowledge during collaboration (Stasser & Titus, 1985) without further help.

On the other hand, as for the *process* aspect of scientific reasoning, the more dyadic members differed from each other on their problem solving strategy, the less they discussed possible intervention or solution plans to alleviate the problem. Furthermore, although dyadic heterogeneity did not seem to have an impact on any other epistemic processes, summing hypothesis generation and evidence evaluation revealed that the more heterogeneous the dyads are the more they engage in "explanatory behavior". The summated results on hypothesizing and evaluating evidence are in accordance with previous findings that dyadic heterogeneity might be beneficial in case of complex problem solving tasks (Bowers et al., 2000; Wiley et al., 2013), and they might explain why more heterogeneous dyads focused less on the solutions: as they spent more "effort" on explaining the problem. The patterns of these findings reflect on the individual vs group differences in scientific reasoning processes. This serves as a strong (although limited: see below) evidence that problem-solving scripts might indeed moderate group processes (Fischer et al., 2013). Thus, diverse problem solving approaches of future professionals can be expected to result in more reflective (explanatory and evaluative) practices.

There are, of course, further explanations for the phenomenon that dyadic diversity seemed to have an impact on the *process* aspect of scientific reasoning while it did not seem to affect the *content* aspect of it. Studies indicate that these two aspects (i.e., content and process) of reasoning might be rather diverse constructs (Mullins et al., 2011; Rittle-Johnson et al., 2001; Zimmerman, 2000). It is therefore possible that, while dyadic heterogeneity measured as *process* knowledge (problem-solving scripts) affect the analogous (i.e., the process) aspect of scientific reasoning, it may have a rather small impact on its non-analogous (i.e., the content) aspect.

7.9.2. Limitations and conclusions

The present study investigated the impact of collaboration and dyadic heterogeneity on pre-service teachers' engagement in scientific reasoning while solving educational problems. We meaningfully differentiated between process and content aspects of scientific reasoning (e.g., Zimmerman, 2000) to capture the different effects of collaboration and heterogeneity.

Generally speaking, our findings demonstrated that collaboration can be useful for scientific reasoning not only for solving scientific problems (Okada & Simon, 1997) but also for future professionals solving professional problems (Fischer et al., 2013). Yet, depending on the professional skills considered to be acquired, curricula may need to consider further educational support. A task where future professionals (pre-service teachers) need to collaborate can be beneficial for engaging them in developing hypotheses or evaluating evidence (Klahr & Dunbar, 1988; Okada & Simon, 1997). Also, they may be more likely to draw conclusions together than individually. These results suggest that situating pre-service teachers in a collaborative problem solving context can help them to practice the skills they might need to become reflective practitioners. On the other hand, without further instructions this increased reflectivity might not guarantee at the same time an increased reference to scientific content. In addition, dyads may consider potential solutions to the problem to a lesser extent than individual reasoners.

These results have further implications how to set up collaborative learning sessions during the curricula. First, dyadic reasoning seems to spontaneously develop in a positive direction regarding engaging in generating hypotheses and evaluating evidence. Therefore, we suggest that in a first phase (exploratory phase), groups may work on their own without further instructions in order to develop explanations and consider supporting / contradicting evidence. After this initial phase groups may receive scaffolds for ensuring the quality of reasoning by referring to scientific content (content quality phase), e.g., by asking groups to revise their explanations and the evidence they used so as it considers to a greater extent scientific concepts, theories and evidence. Third, groups can be prompted to develop solution plans considering at least those explanations that they find the most relevant for explaining the problem (solution phase). In a final stage, groups could be asked to draw conclusions by developing arguments about the limitations of their explanations and solutions regarding the evidence they have and whether further evidence generation would be necessary to make sure they can solve the problem (conclusion phase). Such partially scaffolded phases of collaborative problem solving could potentially bring a great advantage for scientific reasoning compared to individual problem solving. However, further empirical research could clarify to what extent these phases are useful and what could be an optimal level (e.g., Fischer et al., 2013) of scaffolding groups.

One of the main limitation of our study is that we did not measure prior content knowledge; therefore our conclusions regarding the results on the individual advantage on referring to scientific theories and evidence is limited: we cannot claim it with certainty that it results from dyads' focus on shared knowledge vs unique knowledge, neither that such focus emerges from an increased need of knowledge coordination. Future studies therefore should more systematically investigate the impact of cognitive and social factors during collaborative reasoning on the varying outcomes of scientific reasoning content and processes.

A main novelty of our study was to empirically demonstrate how dyadic heterogeneity regarding its members' problem-solving scripts can affect the epistemic behavior of dyads during solving a problem. The results suggest that evidence-based teaching might benefit from heterogeneous group constellation of pre-service teachers. On the other hand, pre-service teachers with diverse problem-solving approaches might need further instructional help to build on that advantage and generate solutions based on the hypotheses and evidence they have developed. Yet, more studies are necessary to understand in detail the effect of group heterogeneity on pre-service teachers' scientific reasoning processes. For example, while our results may be generalizable to co-teaching, bigger teams of pre-service teachers might be even more (or less) affected by prior process knowledge on inquiry.

Finally, the results show predictive validity of problem-solving script measurement in case of the process aspect of scientific reasoning. Similarly, the nonsignificant results with the content aspect may indicate divergent validity if we consider that process and content aspects are considered as diverse (yet, not necessarily unrelated) constructs (Rittle-Johnson et al., 2001; Zimmerman, 2000). Still, further studies are essential to be able to assess the construct validity (regarding convergent and divergent validity) of problem solving scripts.

8. Study 2: Collaborative and Individual Scientific Reasoning of Pre-Service Teachers: New Insights through Epistemic Network Analysis (ENA)[1]

8.1. Abstract

When assessing scientific reasoning both (1) modeling connections in the discourse and (2) doing so at an appropriate grain size can be challenging for researchers. Our study suggests combining a novel theoretical (Fischer et al., 2014) and a novel methodological (Shaffer et al., 2006) framework to respond to these challenges by detecting epistemic networks of scientific reasoning processes in the context of collaborative vs individual problem solving of pre-service teachers. We investigated (1) whether the combination of these frameworks can be fruitfully applied to model scientific reasoning processes and (2) what unit of analysis researchers or instructors should choose to answer questions of interest. One novel aspect of our study is that we compared epistemic networks in case of collaborative vs individual reasoning processes. Our results show that (1) epistemic networks of scientific reasoning can reliably capture reasoning processes when comparing collaborative vs individual reasoning; and (2) propositional and potentially larger units might be considered as "optimal" units of analysis to detect such differences.

Keywords: collaborative problem solving, epistemic network analysis, scientific reasoning

---

[1] This chapter has been developed in co-authorship with Brendan Eagan (University of Wisconsin-Madison), David Shaffer (University of Wisconsin-Madison), Ingo Kollar (University of Augsburg) and Frank Fischer (Ludwig Maximilian University of Munich). Andras Csanadi as a first author of the text (1) has taken major part at every stage of its development and as a primary author (2) takes responsibility for it.

8.2. Introduction

Assessment of scientific reasoning in process data is a critical for the development of appropriate learning support. Although many fruitful approaches have been developed for the evaluation of reasoning and argumentation (Brown, Furtak, Timms, Nagashima, & Wilson, 2010); general theoretical and methodological frameworks that allow analysis of scientific reasoning patterns on multiple layers (e.g., Chi, 1997) are scarce. Consequently, the selection of grain size at an early stage of the analysis and a resulting dilemma surrounding creation of larger units that allow further interpretation of the data (e.g., Weinberger & Fischer, 2006) often limit the generalizability of findings (Chi, 1997; Stegmann & Fischer, 2011). Also, using a pre-defined selection of a unit of analysis might cause difficulties when a researcher or a tutor would like to be more conclusive about the reasoning processes: simultaneously making qualitative and quantitative assessments. For example, a researcher (or tutor) may be interested in ideas, or codes, at a very fine grained (e.g., propositional) level in order to detect "elementary" units of reasoning processes. Meanwhile, she might be also interested in the connections, or relationships, between these ideas or codes captured at that fine-grained level, in order to assess the quality of reasoning processes (Chi, 1997; Weinberger & Fischer, 2006). Moreover, when aggregating data into larger chunks, what would be an optimal choice? Would combining multiple propositions or defining a larger, e.g. sentence units, lead to better representation of reasoning processes? The present study investigates whether a combination of a novel theoretical framework on scientific reasoning (Fischer et al., 2014) as well as a novel methodological approach on modelling reasoners' epistemic networks (Shaffer, 2006) can be meaningfully combined 1) to analyze patterns (epistemic networks) of scientific reasoning and 2) to disambiguate the question on grain size selection and data aggregation when assessing patterns (epistemic networks) of scientific reasoning.

8.3. Scientific reasoning and argumentation

There are different theoretical frameworks to conceptualize and analyze scientific reasoning. Many follow a "structural" approach, focusing on the structure of argumentation (see Brown et al., 2010) while others emphasize the role of engagement in scientific reasoning processes (Okada & Simon, 1997). Our work belongs to the latter stream of research understanding scientific reasoning as engagement of individuals or groups in a sequence of epistemic processes (Fischer et al., 2013). According to this model, scientific reasoning involves reasoners identifying an existing problem (Problem identification), articulating questions of how to proceed with their reasoning processes (Questioning), derive possible explanations of the problem (Hypothesis generation), construct artifacts, such as intervention plans, to solve the problem (Generating solutions), generate and collect information (Evidence generation), evaluate that information (Evidence evaluation), engage others in the reasoning process (Communicating & scrutinizing), and draw conclusions (Drawing conclusions). Earlier studies found that both individual and collaborative reasoning in a professional problem solving context can be reliably coded using this framework (Csanadi, Kollar & Fischer, 2016).

8.4. Collaborative vs. individual scientific reasoning processes

Collaborative scientific reasoning has the potential to lead individuals to higher engagement in epistemic processes such as hypothesis generation and evidence evaluation compared to reasoning alone (Okada & Simon, 1997; Teasley, 1995). Similarly, more recent findings (Csanadi et al., 2016) showed that when pre-service teachers solved a problem from their future practice as dyads, they engaged more in hypothesis generation (i.e., trying to find an explanation to the problem) but less in generating solutions than individuals did. Nevertheless, this purely frequency-based approach for analysis to count the occurrence of certain codes has clear constraints. Most importantly, it cannot be conclusive enough regarding the patterns of epistemic processes that can characterize collaborative vs individual reasoning. For example, although dyads were found to be more explanatory, indicated by a higher engagement in hypothesizing, whether they did this in a more evidence-based manner (i.e. if they made more connections between hypothesizing and evaluating evidence) remained unclear. Being able to identify such connections or patterns in the data is, therefore, important for assessing quality aspects of scientific reasoning.

8.5. Selection of grain size and data aggregation to capture patterns of reasoning

To assess and compare reasoners with respect to the patterns of the epistemic processes they engage in, researchers should find answers to two related questions. First, what is an appropriate grain size (i.e., unit of analysis) and second, how should coded data be aggregated in order to gain a deeper understanding of the quality and features of the reasoning processes. Many researchers emphasize that data segmentation should be a separate and preceding step to coding (Chi, 1997; Strijbos, Martens, Prins & Jochems, 2006). This would mean that the division of verbal data into chunks that carry meaningful information for further analysis should precede further analyses. However, this early selection of the unit of analysis has its limitations (e.g., Chi, 1997). Especially the use of smaller grain sizes (e.g., propositional unit) allow for a more fine-grained analysis of reasoning processes (e.g., to interpret the relation between independent clauses of compound sentences) and allow for frequency-based analyses. Indeed, many quantitative approaches to the analysis of scientific reasoning processes (e.g., Okada & Simon, 1997) suggest analyzing frequencies of single categories. However, considering that discourse moves are not unrelated to each other, relying on solely frequency-based information of data can lead to missing meaningful patterns of discourse (Cress & Hesse, 2013). At this point an emerging concern of data aggregation (Stegmann &

Fischer, 2011), i.e., how the researcher/tutor can make higher level inferences based on data coded at a lower grain size, often generates uncertainty. When looking for relationships between coded units (e.g., propositions), how far these units can fall from each other? Can we meaningfully detect relationships between two neighboring units or does allowing for slightly "longer distance" connections increase explanatory power? A method that allows more adaptable choice of grain size (Siebert-Evenstone et al., 2016), such as considering multiple units of analysis instead of relying on a pre-defined selection in order to model scientific reasoning could help to answer such questions.

Another issue associated with coding-independent segmentation may arise if some codes turn out to be highly frequent ones while others occur relatively rarely. "Uneven" frequency distributions can bias further analyses of the dataset (e.g., Csanadi, Daxenberger, Ghanem, Kollar, Fischer & Gurevych, 2016). For example, high frequency codes might generate many connections with each other while also being related to many other codes. On the other hand, low frequency codes may lack enough connections with other codes to demonstrate the power to discriminate between epistemic networks of different groups (e.g., dyads vs individuals). Thus, in case of modeling reasoning processes, this can mean that some reasoning patterns may emerge as mere artifacts while other connections in the data may remain undetected, and therefore, models of scientific reasoning should account for such limitations.

To summarize, using a hierarchical segmentation procedure and reliance on solely frequency-related information when analyzing scientific reasoning processes and comparing reasoners, leaves open the questions of (1) how to aggregate and identify meaningful larger patterns in the data that can (2) help more validly capture the reasoning performance beyond simply counting the occurrences of single codes.

8.6. Epistemic Network Analysis: A Method to Analyze (Multiple Scopes of) Scientific Reasoning

One solution of the abovementioned problems can be to code on multiple levels of granularity (Stegmann & Fischer, 2011). As Chi (1997) notes, this approach has the advantage of leading to more reliable results and interpretations at different levels. Generally speaking, segmentation might be a matter of the researchers' focus of interest (Chi, 1997), the theoretical framework they apply (Clara & Mauri, 2010), the nature of data (e.g. synchronous vs asynchronous discussions) and more. Still, selecting multiple levels of analysis can contribute to more valid interpretations about the data (Chi, 1997; Weinberger & Fischer, 2006) as different lenses may capture different aspects of collaborative learning and reasoning processes.

Epistemic Network Analysis (ENA; Shaffer, 2006) is a method to identify meaningful and quantifiable patterns in discourse/reasoning. It can provide an alternative to the widespread "code and count" approach. ENA moves beyond the traditional frequency-based assessments by examining the structure of the co-occurrence, or connections in coded data. Moreover, compared to other methodological approaches, e.g., sequential analysis (see in Cress & Hesse, 2013), ENA has the novelty of (1) modeling whole networks of connections and (2) it affords both quantitative and qualitative comparisons between different network models.

A main theoretical assumption of ENA is that repeated co-occurrences of two or more codes in the discourse can reveal epistemic networks which characterize an underlying Discourse (Gee, 1999; Collier et al., 2016), e.g., to collaborative (vs. individual) scientific reasoning. To identify a unit of analysis for calculating such co-occurrences, ENA provides an adaptable feature: the *moving stanza window size* (MSWS; Siebert-Evenstone et al., 2016).

The term stanza window refers a window or scope within which ENA is searching for connections. This means that a MSWS=1 allows search for connections only between a proposition of reference and its preceding proposition. Therefore, a MSWS=1 results in connections only between neighboring propositions. A MSWS=2, however, allows one further step: it allows connection between a proposition of reference and the two preceding propositions. By changing MSWS from smaller values to larger it is possible to open the "search window" from very narrow context to wider ones. As a result, the researcher or tutor can look for connections not only within propositions (as in case of "coding and counting" approaches) or between neighboring propositions, but even between propositions that are two, three or more steps further from each other in the discourse. In short, it offers the advantage of multiple scopes for analysis. Here we aim to investigate if ENA can reveal some characteristics of collaborative (compared to individual) scientific reasoning processes as well as to articulate what grain sizes should be considered when using ENA for that analysis.

Furthermore, ENA provides the opportunity to quantitatively and qualitatively compare different epistemic network models with each other. Quantitative comparison is possible by using calculated centroids for every epistemic networks generated by ENA. Such centroid values are determined by the strength of connections between nodes in the epistemic network. Nodes are the codes (such as epistemic processes, see below) while the strength of connections between them are generated based on their local co-occurrences (within each stanza window: see above). These centroid values can be used for quantitative analyses. Furthermore, qualitative comparison of epistemic networks is possible using various options for visualization. One option is "Subtracting networks" which means contrasting two network models by subtracting their nodes and connections weights from each other. A resulting "subtracted network" represents the difference between two reasoning networks and therefore, can illustrate what makes dyadic reasoning different from individual reasoning.

8.7. Research questions

RQ1: Do collaborative and individual reasoners exhibit different epistemic networks of scientific reasoning while solving a professional problem?

While earlier studies demonstrated differences between collaborative and individual reasoning in terms of their engagement in different epistemic processes (Csanadi et al., 2016; Okada & Simon, 1997), these results were mainly frequency-based. E.g., the researchers compared proportions as well as raw frequencies of engagement in different epistemic processes, such as evaluating evidence or hypothesizing. Thus, an open question is whether dyads also differ from individuals in the patterns of epistemic processes they engage in during scientific reasoning. In this study we address this question using ENA (Shaffer et. al. 2009) to capture meaningful patterns of co-occurrences between epistemic processes (i.e., epistemic networks of scientific reasoning), and to compare dyads with individual reasoners.

Epistemic networks can, however, also be defined based on larger speech units (e.g., across multiple propositions) and we can also implement larger grain sizes beyond analyzing neighboring propositions or within sentences. To fully answer RQ1, therefore, we investigated whether some grain sizes can provide potentially better explanation of patterns in the data than others.

RQ2: Do the epistemic networks we detect investigating RQ1 differ from epistemic networks based on the same data set that has been randomly resorted (i.e. with the same frequency information)?

ENA models co-occurrences of codes, since some codes occur more frequently than others, it is more likely that these highly frequent codes make connections (co-occur) with other codes more often than lower frequency codes. Consequently, ENA may "overestimate" some connections. Therefore, to answer our second research question, we compared ENA results from RQ1 to ENA results obtained from a dataset that contained only frequency information of the original discourse (see below). If the epistemic networks identified in relation to RQ1 cannot be explained merely by the frequency distribution of epistemic processes, the epistemic networks detected in relation to RQ1 should differ from the epistemic networks of the randomly resorted dataset.

8.8. Method

The data analyzed in this study is a re-analysis of process data from another study (Csanadi et al., 2016). In the original study N=76 preservice teachers (59 female, M_{Age}=21.22, SD_{Age}=3.98) solved a problem case from their future profession in one of two between-subject conditions: either as individuals (N=16) or as dyads (N=30 dyads). Think aloud and discourse data of their problem solving were first manually segmented into propositional units and then coded for further analysis. The coding scheme of that study was developed based on the framework of scientific reasoning by Fischer et al. (2014). Epistemic processes identified by the framework (see above) were applied (Table 1): Problem identification for an initial attempt to build an understanding of the problem; Questioning for statements or questions triggering further inquiry; Hypothesis generation for developing explanations of the problem; Evidence generation for reference to information or lack of information that could support a claim; Evidence Evaluation to evaluate a claim; Communicating and scrutinizing for planned discussions with others (e.g., in order to find out further information); Drawing conclusions for concluding outcomes of reasoning. Finally, the epistemic process of "Constructing artefacts" (in Fischer et al., 2014) was operationalized as developing interventions or solution plans, and such propositions were labelled as Generating solutions. Moreover, the codes for

Evidence generation and Evidence evaluation were merged into Evidence evaluation. Both segmentation (79.73% of agreement by Coder 1 and 85.09% of agreement by Coder 2) and coding (κ = 0.68) proved to be reliable. We used this dataset (original dataset) to analyze further in our present study.

We used the abovementioned original dataset to answer RQ1. To be able to answer RQ2 we created a randomized dataset in the following way. Using the original dataset within each dyad and individual participants we created a random sequence of the pre-segmented propositions (Csanadi et al., 2016). That meant, the original sequence of propositions were randomized while the relative frequency of propositions was preserved (no propositions were deleted). This new randomized dataset preserved the information of the occurrence of epistemic processes, yet, in a randomized order; containing the information to which individual or dyad the epistemic processes belong to, how frequently they occur, but without any information regarding their sequence in the original dataset.

We used ENA to identify epistemic networks of scientific reasoning in order to answer both RQ1 and RQ2. We built epistemic network models using ENA in four steps. First, we calculated co-occurrences between epistemic processes (MSWS=1, means rotation was applied) for dyads and for individuals. At the same time ENA automatically generated a centroid value for each dyad or individual that served as a numeric representation of their epistemic network and it was included in further analysis to compare dyadic and individual epistemic networks of scientific reasoning. Second, mean, or “average,” networks were defined for both the dyadic and the individual reasoning conditions, respectively. Each of these networks visually represented all the connections that participants (dyads or individuals) generated in the given condition. Third, we quantitatively compared epistemic networks for dyads with epistemic networks for individuals by comparing the mean centroid values (calculated in step 1) in the two conditions. Fourth, we subtracted the mean dyadic and mean individual networks from each other (by using the "Subtracting networks" option in ENA). The resulting subtracted networks visualized what connections contributed to the difference between the two reasoning conditions (dyadic vs individual, calculated in step 3).

To be able to fully answer RQ1 regarding grain size, we sequentially set MSWS from 1 to 7, step-by step, performing the same analysis for each stanza window size. The resulting epistemic network models at each MSWS level allowed us quantitative as well as qualitative (visual) comparisons.

To answer RQ2, we used the randomized dataset selecting the same parameters and performing the same analysis as in case of RQ1. We compared the outcomes of this analysis with the ENA results from RQ1.

8.9. Results

RQ 1: To answer RQ1, as a first step, we compared dyadic and individual networks at the grain size of MSWS=1 which lead to the following results. The mean centroid value for individuals' epistemic networks (M=.21,SD=.32) was significantly different from the mean centroid value for dyads' epistemic networks (M=-.11,SD=.21), $t(44)$=3.65, p<.01, d=1.32. Plotting epistemic networks (Figure 8) further revealed that the central epistemic process accounting for most of the connections was evidence evaluation. Moreover, in case of dyads evidence evaluation showed more complex network than in case of individuals: for dyads it was connected to hypothesis generation, communicating and scrutinizing, generating solutions and non-epistemic propositions; while in the case of individuals it was only connected to hypothesis generation and generating solutions. Finally, subtracting individual from dyadic networks revealed that in case of individual networks it was solution generation rather than evidence evaluation that played a central role in contrast to dyadic networks where only evidence evaluation showed multiple connections after subtraction.

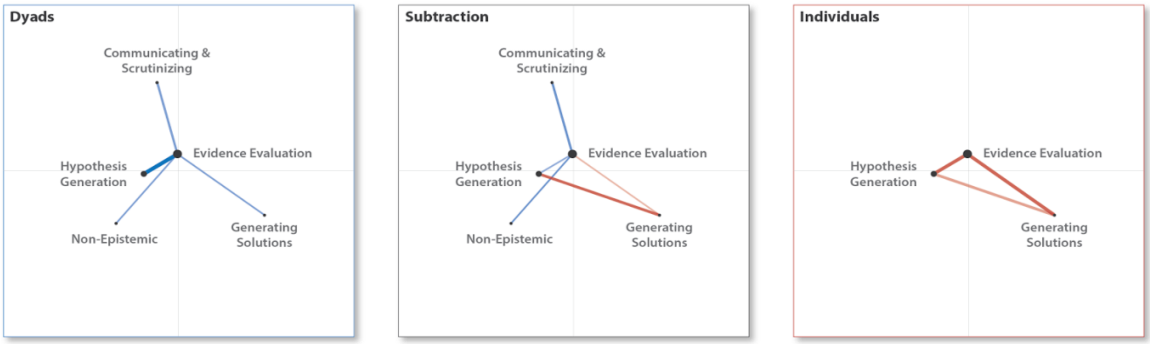

Figure 8. Epistemic networks of dyads (blue, left), individuals (red, right) and the difference between their networks (center) using the original dataset.

To completely answer RQ1 and in order to see whether there is an optimal grain size that can best capture the differences between epistemic networks of dyads and individuals, we compared epistemic networks at $1 \leq$ MSWS ≤ 7 levels which led to the following results. All comparisons were statistically significant at least under $p<.01$. Although effect size showed a small increase at every MSWS level, these differences were small: the explained variance increased only by 5.35% (ΔR^2=.05) from MSWS=1 (R^2=.30) to MSWS=7 (R^2=.36). Finally, a visual inspection of the epistemic networks conducted at $1 \leq$ MSWS ≤ 7 levels suggested highly similar patterns at every MSWS levels (see Figure 8).

RQ 2: Similar to the outcomes of RQ1, when using the randomized dataset, the mean centroid value for individuals' epistemic networks (M=.17,SD=.26) was significantly different from the mean centroid value for dyads' epistemic networks (M=-.09,SD=.20), $t(44)$=3.35, p<.01, 95%, d=1.15. Plotting epistemic networks (Figure 9), however, revealed no visible difference between dyadic and individual networks. Dyadic and individual networks showed identical patterns regarding complexity: connections occurred among the three most frequent epistemic processes: hypothesis generation, solution generation and evidence evaluation. This was in clear contrast with the results of RQ1 where epistemic networks were different for collaborative vs individual reasoning (Figure 8). A further important difference is that Figure 9 does not indicate any central epistemic process, neither for dyadic and individual nor for the subtracted pattern. Moreover, Figure 9 shows very low level of network complexity for dyads (connections among the highest-frequency activities) compared to Figure 8. Finally, the subtracted network model on Figure 9 consists of only blue lines, indicating that dyads made more connections among the highly frequent codes than individuals.

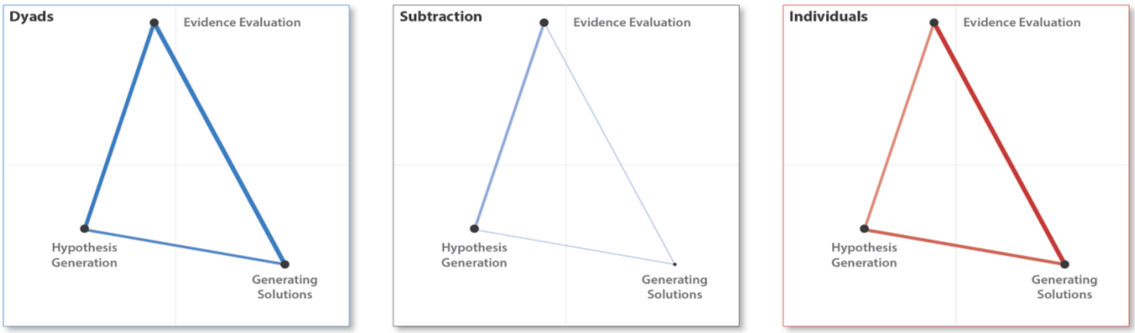

Figure 9. Epistemic networks of dyads (blue, left), individuals (red, right) and the difference between their networks (center) using the randomized dataset.

8.10. Discussion

The two main aims of our study were (1) to see whether we can aggregate data to capture meaningful patterns (epistemic networks) of scientific reasoning processes regarding collaborative and individual reasoning (RQ1 & RQ2) and (2) to search for an optimal grain size, or unit of analysis, for such aggregation (RQ1). We sought to answer these questions by the application of a novel theoretical framework on scientific reasoning (Fischer et al., 2014) and a novel methodological approach on modelling epistemic networks (Shaffer, 2006).

The outcomes for RQ1 suggest that epistemic networks of scientific reasoning can meaningfully differentiate between collaborative and individual reasoning processes. More specifically, dyads seemed to engage in a more complex manner in scientific reasoning compared to individuals: they made more connections between epistemic processes (specifically, with evidence evaluation). Moreover, while individual reasoning was rather solution-focused; dyadic reasoning seemed to be more evidence-focused. These results are also in accordance with previous frequency-based findings (Csanadi et al., 2016; Okada & Simon, 1997).

To be able to fully answer RQ1 we ran further analyses at different stanza window sizes that resulted in patterns quite similar to those in Figure 8. On the one hand, this suggests the robustness of our findings, on the other, a question of the optimal grain size to detect meaningful patterns of scientific reasoning cannot be conclusively answered. A partial answer is, however, that choosing larger speech unit (e.g., sentences) at a first step may represent reasoning patterns in the data at least closely as well as propositions do. Yet, further empirical research could test (1) whether this is true and if (2) varying stanza window sizes on sentence units would lead to different results. Based on the results of this study and considering the exhaustiveness of hand-coding procedure, however, choosing larger units of analysis that still carry the information needed to model scientific reasoning may be an efficient choice for the researcher/tutor.

The outcomes on RQ2 show that epistemic networks extracted on discourse data (original dataset) are likely to be valid models for the evaluation of reasoning patterns in the data as they are not reducible to the frequency distribution of codes. Furthermore, it is clear that merely frequency-information in the data resulted in only “poor” network models: networks represented solely the most frequent codes and their connections. Additionally, after subtracting those networks the results suggested that dyads made more connections everywhere. These results did not add much explanatory value to the frequency-based outcomes of the earlier findings (Csanadi et al., 2016), which underlines the assumption that ENA conducted on real discourse data can detect meaningful patterns of scientific reasoning.

Finally, the results imply that identifying epistemic processes on the propositional level and aggregating data by conducting epistemic network analysis can offer a powerful way to meaningfully assess scientific reasoning in discourse.

8.11. Final conclusions

Our results have further important consequences.

First, the theoretical (Fischer et al., 2014) and the methodological (Shaffer, 2009) frameworks could be fruitfully combined to result in a series of robust analyses of identifying epistemic networks of scientific reasoning.

Second, dyadic vs. individual reasoning networks can be valid models of scientific reasoning in discourse. Yet, we need more empirical research to see if this result holds as well as see the predictive validity of our findings. For example, the extent to which dyads' more extensive connections could potentially predict learning outcomes and whether some connections might play a stronger moderating role in that process, are questions for future research.

Finally, additional analyses that can more directly address the impact of frequency distribution of codes on epistemic networks could also contribute to conclusions regarding the validity of the findings. For example, alternative measures provided by ENA could account for "imbalanced" frequency distribution in the data. Those measures could apply, for example, some weighting method for assigning less weight to higher frequency codes or to connections among higher frequency codes, in order to reduce the chance of detecting artefactual connections due to higher probability of co-occurrence between high-frequency codes. Similarly, if ENA could generate a simple frequency-based epistemic network model (similar to the outcomes on RQ2) and would allow its subtraction from the epistemic network model on the real dataset; that would afford the visualization of reasoning patterns beyond highly frequent connections. Yet, such measures should be implemented with caution: connections captured in the discourse should always represent connections in the Discourse (Gee, 1999).

9. General discussion

This chapter is concerned with the question whether the main aim of the dissertation, to find out whether collaboration can be a beneficial context to engage future practitioners to solve professional problems as scientifically knowledgeable practitioners, has been fulfilled. With respect to that concern, this chapter discusses how the research questions (Chapter 7) can be answered on the basis of the results of the two empirical studies conducted (Chapter 8 and Chapter 9), and accounts for the theoretical, methodological and practical implications of those results.

9.1. Summary of the Studies

The present dissertation aims to answer the question if engaging future practitioners to collaborate with each other while solving authentic problems from their future professional practice would be beneficial for them to solve problems in a scientifically knowledgeable manner, i.e., to reason more scientifically compared to individual problem solvers. In order to be able to answer that question the present work accounts for two related sub-problems. First, this dissertation argues that when comparing groups of reasoners with individual reasoners group composition (i.e., within-group heterogeneity regarding their members' problem solving scripts) might moderate the findings. Second, this dissertation claims that the mere application of a "coding and counting" approach is not sufficient enough for measuring scientific reasoning and, a complementary methodological approach, i.e., Epistemic Network Analysis (ENA) can bring the potential to reveal a more detailed picture on the process aspect of scientific reasoning compared to those "coding-and-counting" approaches.

Studies suggest that when it comes to solving professional problems, (future) practitioners such as pre-service teachers often do not reason like scientifically knowledgeable practitioners do such as they do not apply scientific knowledge in a systematic manner (Hetmanek et al, 2015; Yeh & Santagata, 2015). It indicates that future practitioners have difficulties with both the content and the process aspects of scientific reasoning. It is a reasonable assumption therefore, that asking them to collaborate while solving problems might foster their engagement in scientific reasoning (Okada & Simon, 1997; Rummel & Spada, 2005). Yet, so far there is a lack of empirical investigations in this area. Similarly, although heterogeneous group composition is often reported (Bowers et al., 2000; Canham et al., 2012) to bring benefits to problem solving; no studies have investigated so far whether group heterogeneity regarding prior knowledge on problem solving, i.e., problem solving scripts, may facilitate engagement in scientific reasoning. Finally, although previous literature showed that the exclusive application of the method of “coding and counting” is not an optimal way to analyze reasoning processes (e.g., Jeong, 2005), and that studies should focus on the relationship or patterns between epistemic processes or “skills” constituting scientific reasoning (Klahr & Dunbar, 1988; Suthers, 2005), previous approaches seem to be limited with respect to the analysis of patterns (i.e., networks) of scientific reasoning.

To fill the abovementioned research gaps, two studies (using the same data) with pre-service teachers have been conducted. Both investigated the main question to what extent collaboration is a beneficial context to solve problems as scientifically knowledgeable practitioners. In addition, Study 1 (Chapter 8) investigated the effect of group heterogeneity regarding problem solving scripts on the engagement in scientific reasoning while Study 2 (Chapter 9) investigated if epistemic network analysis can be an appropriate methodological approach to assess scientific reasoning. Additionally, the present work differentiated between two aspects of scientific reasoning: its content and process aspects. The following subchapters summarize the results of each study.

9.1.1. Summary of Study 1

The first experimental study found that asking pre-service teachers to collaborate during solving educational problems has divergent impact on the process and the content aspects of scientific reasoning. To be more specific, collaboration fostered the engagement in scientific reasoning *processes* such as hypothesis generation; evidence evaluation and drawing conclusions. On the other hand, it decreased the engagement in generating solutions to the problem. Regarding the *content* aspect, dyads applied proportionally less scientific content than individuals did.

When investigating the effect of heterogeneity of group members' problem solving scripts on the engagement in scientific reasoning, the results showed that the more heterogeneous the dyads were, the more they engaged in the *processes* of hypothesis generation and evidence evaluation altogether. At the same time, the more heterogeneous the dyads were the less they engaged in generating solutions to the problem. Dyadic heterogeneity did not seem to affect the content aspect of scientific reasoning in this study.

9.1.2. Summary of Study 2

The findings of the second study indicated that patterns of epistemic processes of scientific reasoning can be captured by applying the method of epistemic network analysis (ENA). Moreover, such patterns could differentiate between collaborative and individual reasoners. Network subtractions showed that the connections of evidence evaluation with hypothesis generation, communicating and scrutinizing as well as with non-epistemic propositions distinctively characterized collaborative reasoning, while the connections of solution generation with hypothesis generation and evidence evaluation were characteristic for individual reasoning. Finally, these connections appeared to represent the co-occurrences between epistemic processes in a more interpretable manner compared to when the sequence of epistemic processes was randomized in each of the discourse and think aloud.

9.2. Conclusions and Implications of the Results

9.2.1. Conclusions of Study 1

One of the main finding of Study 1 was that collaboration affected the process and content aspects of scientific reasoning differently. Reasoners' engagement in epistemic processes of hypothesis generation, evaluating evidence and drawing conclusions showed increase in case of collaborating partners compared to individual reasoners, while solution generation decreased as an effect of collaboration. On the other hand, individuals referred to scientific content more often than collaborating partners did.

Regarding the main research question whether collaboration can be an beneficial context for engaging future practitioners in scientific reasoning, one of the main findings was that the collaborative context showed positive effect for the most part of the *process* aspect: it increased the engagement in epistemic processes such as hypothesis generation, evidence evaluation and drawing conclusions, but decreased engagement in generating solutions. The earlier assumptions of this dissertation indicated that a collaborative context might be authentic for scientific reasoning (Simon et al., 1981; Osborne, 2010) and that it might lead to communicational affordances such as engagement in interactive reasoning (Chi, 2009) as the different reasoning partners can complete and challenge each other's thoughts (Dunbar, 1995) as well as the distribution of cognitive load that may otherwise impede individual problem solving processes (e.g., Kirschner et al., 2011). It is possible that developing explanations (i.e., generating hypotheses) and evaluating evidence are those epistemic processes that can benefit the most from a collaborative reasoning context, because they might require more interactivity in the form of questioning or challenging each other's ideas compared to e.g., generating solutions. Further studies could potentially investigate this assumption by, for example, comparing the level of interactivity (Chi, 2009; Teasley, 1997) and argumentativeness (Asterhan & Schwarz, 2009) in case of hypothesis generation and solution generation. Another potential is that solving an authentic problem might have increased cognitive load in case of individuals but not for groups, because dyadic members could rely on each other's reasoning (e.g., Kirschner et al., 2011). Yet, as the present study did not measure cognitive load, this explanation remains speculative. Nevertheless, the results are in accordance with the findings of earlier studies (e.g., Okada and Simon, 1997) with respect to hypothesizing and generating/evaluating evidence which indicates the generalizability of analyzing these epistemic processes in non-scientific, i.e., practice-oriented domains. The similar pattern of the results compared to earlier findings (Okada & Simon, 1997; Teasley, 1995) can also serve as an argument for the content validity of the coding instrument applied in this study, at least with respect to the epistemic processes of hypothesis generation and evidence evaluation. Moreover, the results serve as one of the first empirical demonstrations on how epistemic processes may occur in the "practitioners' mode" of scientific reasoning and what skills might constitute "scientifically knowledgeable practice" (Fischer et al., 2014). Yet, further studies should be conducted in other non-scientific domains to test for the domain-generality of scientific reasoning processes within practical domains as well as to be able to draw an overall conclusion that the same epistemic processes apply for non-scientific as for scientific domains.

A further main finding regarding the process aspect of scientific reasoning was that collaboration affected solution generation in a negative manner. This result taken together with the just discussed positive effect on the other epistemic processes seems to represent a trade-off between epistemic processes that are often understood as core to scientific inquiry such as hypothesizing and evidence evaluation (Klahr & Dunbar, 1988; Okada & Simon, 1997), and the epistemic process of rather practical characteristics such as generating solutions. This might suggest that while collaboration was characterized more by an explanatory-investigative epistemic focus, individual reasoning was more solution-oriented without necessarily elaborating on the nature of the problem. One possible explanation can be the idea-coordination between reasoning partners (Barron, 2000; Roschelle & Teasley, 1997; Okada & Simon, 1997) might account for the effect of collaboration on the increased engagement in explanatory-investigative processes. When solving a problem collaborating partners may need to coordinate their ideas in order to develop a common understanding about the nature of the problem (e.g., Roschelle & Teasley, 1995). This coordination within dyads might lead to longer discussions on the potential explanations of the problem (hypothesis generation) and may argue for or against their assumptions based on evidence (evaluating evidence). As a consequence of this increased engagement in explanatory-investigative epistemic processes, dyads might sense it rather superfluous to engage in other epistemic processes such as solution generation. Yet, it is somewhat contradictory to this interpretation that dyads were more likely to draw conclusions than individuals were. Another potential explanation is that the overall process of solving complex problems might result in higher cognitive load for individuals who are novice problem solvers in their field (i.e., *pre-service* teachers) compared to groups of novices (e.g., Kirschner et al., 2011). For novice problem solvers it might be cognitively effortful to think about different aspects of solving the problem (e.g., how to progress with solving the problem, what scientific theory may be relevant to apply). While dyads have further resources, i.e., a reasoning partner, to cope with cognitively effortful questions by relying on the partner's knowledge and contribution to the discourse (Hollingshead & Brandon, 2003; Kirschner et al., 2011; Pea, 2003), for individual reasoning there is no such scaffolding opportunity to cope with an increased cognitive load. This lack of necessary cognitive resources might result in an epistemic focus on staying pragmatic and thinking about solutions more than considering potential explanations of the problem (generating hypotheses) and connecting those explanations to evidence (evaluating evidence). An interesting question to find out would be if individuals build their solutions based on systematic inquiry of potential explanations about the problem or if they are more prone to come up with solutions without building them on earlier explanations of the problem. Earlier studies (Chi, 2009; Dunbar, 1995; Okada & Simon, 1997; Teasley, 1997) indicate that the communicational affordance of collaborative (scientific) reasoning might lead to more reflection and an increased need between collaborating partners to explain upcoming ideas, e.g., for clarification, compared to individual reasoning where such communicational challenges do not exist. Further studies might try to find an answer to the question to what extent coordination attempts of dyads or lack of (self-)reflective reasoning of individuals may account for the trade-off effect between explanatory-justification processes and generating solutions.

Besides the abovementioned effects of collaboration on the process aspect of scientific reasoning, collaborative reasoning seems to affect the *content* aspect in a negative manner. Dyads applied proportionally less scientific content during problem solving than individuals did. It is notable that scientific reasoning did not benefit from collaboration with respect to the application of scientific content. One potential explanation for this result is that although different learning partners might have brought unique ideas (i.e., scientific knowledge) into the discussion, they may have focused on mutually shared knowledge (Stasser & Titus, 1985), for instance to reach and keep up a mutual understanding on the problem (Roschelle & Teasley, 1995). Yet, it is somewhat speculative to draw such conclusions considering that no prior scientific content knowledge was measured in this study. Thus, there is no information about the extent to which reasoning partners initiated their original ideas or to what extent they focused on shared knowledge. Another possible explanation is a more methodological one. Namely, that the coding scheme on scientific content use did not measure scientific content at the appropriate level. Specifically, although the applied coding scheme was able to differentiate between different scientific content areas (e.g., memory/learning; motivation; anxiety) in a reliable manner, these content areas were analyzed altogether without an even finer differentiation between content coding. For example, a lower level coding of which scientific theories reasoners refer to might be informative about whether collaborating partners might have referred to, e.g., more unique content than individuals did. Similarly, addressing the "appropriateness" or the quality of the reference to scientific content might also reveal potential benefits of groups over individuals. Therefore, by applying a rather "low threshold" for the operationalization and the measurement of what would count as a reference to scientific content (i.e., identifying and then merging scientific content areas) might not be the (only) valid way to capture the content aspect of scientific reasoning, i.e., the application of scientific knowledge. At any rate, future studies should measure prior knowledge on scientific content in order to be able to investigate whether a lack of sharing unique knowledge is indeed an issue for collaborative scientific reasoning about professional problems (Stasser & Titus, 1985).

A further novelty of the present study is the empirical demonstration of how the level of group heterogeneity regarding dyadic members' prior knowledge on problem solving processes (i.e., problem solving scripts) affects the engagement of dyads in scientific reasoning. The results on the effect of dyadic heterogeneity with respect to dyadic members' problem solving scripts on the engagement in epistemic processes of scientific reasoning demonstrated a similar pattern as the comparison of collaborative vs individual reasoning. Namely, the more heterogeneous the dyads were, the more they showed the explanatory-justification focus of collaborative reasoning: i.e., higher engagement in hypothesizing and evidence evaluation and lower engagement in solution generation. The more homogeneous the dyads were, the more they seemed to show the solution-focus that was so characteristic of individuals: i.e., lower engagement in hypothesizing and evidence evaluation and higher engagement in solution generation. These results extend the earlier findings of group heterogeneity (e.g., regarding expertise) on problem solving (Wiley & Jolly, 2003). The present work demonstrated how group heterogeneity regarding prior script-like knowledge (Fischer et al., 2013; Kollar et al., 2007), such as problem solving scripts, affect scientific reasoning processes in a problem solving context. The results, on the one hand, may indicate a stimulation effect (Dunbar, 1995; Paulus, 2000) of dyadic heterogeneity on the engagement in epistemic processes traditionally understood as being prototypical for scientific reasoning (Klahr & Dunbar, 1988; Okada & Simon, 1997). On the other hand, these results may also be either partially or fully explained by an increased coordination demand between heterogeneous reasoning partners (Bullough et al., 2002; Rummel & Spada, 2005). Further analysis should reveal if knowledge co-construction (Chi & Wiley, 2014; Roschelle, 1992) is affected by dyadic heterogeneity and if so, how those processes might mediate the overall extended engagement in hypothesizing and evaluating evidence. For example, do members of heterogeneous dyads show a complementary effect (i.e., complete each other's strategies) by exchanging different perspectives (Paulus, 2000)? Alternatively, do they engage more often in dialectic reasoning and challenge each other's views (Métrailler et al., 2008; Osborne, 2010)? Or do they, after all, spend relatively more effort to coordinate ideas, i.e., asking for and giving clarification (Asterhan & Schwarz, 2009; Okada & Simon, 1997; Teasley, 1997)?

A further explanation of the negative impact of dyadic heterogeneity on the engagement in solution generation might be that, similarly to the discussion of the results of collaborative vs individual reasoning, the extended engagement in explanatory-justifying behavior may account for the result. This would mean that the longer reasoning partners engage in explaining the problem and arguing about it by evaluating evidence the less time or effort they need to engage in generating solutions, because they might have developed an elaborated understanding on the problem through extended earlier discussion. Yet, less time or effort might not be the best indicator of the quality of ideas. Therefore, further studies should test the question if an increased engagement in hypothesizing and evaluating evidence might indeed contribute to the "fluency" of generating solutions and also to the appropriateness or quality of those solutions.

Moreover, although the time for the overall problem solving was not fixed, i.e., reasoners were not stopped to talk, all participants were informed that they had about ten minutes for the problem solving task. This framing of available time might have resulted in a perception that the task needed to be fulfilled during this timeframe. As a result, those pairs who discussed longer explanations and evidence might have spent less time and effort on discussing solutions, because they may have thought they do not have much time for that or because they felt the task completed by their earlier discussion. Further studies should clarify and even systematically test the effect of (perceived) time available for the problem solving task on the engagement in the different epistemic processes.

Furthermore, the first study could not demonstrate a potential impact of dyadic heterogeneity on the content aspect of scientific reasoning. On the one hand, studies report a positive effect of group diversity on complex problem solving (Bowers et al., 2000; Canham et al., 2012) and point out that group diversity might also stimulate scientific reasoning

(Dunbar, 1995). Other studies point to potential difficulties that heterogeneity might bring to exchanging ideas and solving problems in groups such as increased coordination demands (Bullough et al., 2002; Rummel & Spada, 2005) or collaborative inhibition of exchanging information (Hirst & Echterhoff, 2012). On the other hand, dyadic heterogeneity regarding dyadic members' problem solving scripts did not seem to lead to neither stimulation nor inhibition with respect to the application of scientific content knowledge. First, it is possible, that there is no robust effect of dyadic heterogeneity regarding dyadic members' problem solving scripts on the application of scientific content. However, it is also possible that potential confounds or covariates did not allow for the detection of this effect. For example, it is arguable that dyadic heterogeneity regarding prior knowledge on scientific content might explain a certain amount of variance in the application of scientific content. If future analysis accounts for that effect, it is possible that the effect of heterogeneity regarding problem solving scripts become detectable.

In addition, future studies should address the question to what extent dyadic heterogeneity regarding other frequently reported potential mediators of collaborative performance, i.e., internal collaboration scripts (Fischer et al., 2013; Kollar et al., 2007) influence the results. Studies indicate that the knowledge about how to collaborate with a reasoning partner (i.e., how to engage in productive discourse) might have an impact on the extent to which reasoning partners may engage in such discussions (Fischer et al., 2013; Vogel et al., 2016). The relationship between collaboration scripts which guide the social aspect of collaboration (Fischer et al., 2013) and problem solving scripts which guide the epistemic aspect of collaboration (Weinberger et al., 2005) should be clarified: first, to what extent they are divergent constructs, and second, whether they have any interaction effect on how future professionals engage in scientific reasoning.

Besides, the results that (1) prior process knowledge affected the process aspect of scientific reasoning and that (2) the more similar (homogeneous) group members were regarding their problem solving scripts the more they reasoned like one (individual) might also serve as an empirical argument for the validity of the problem solving script measurement tool in the first study which was developed to capture individual differences in problem solving scripts.

Finally, the finding that collaboration affected the process and content aspects of scientific reasoning in a different way suggests that scientific reasoning might not be a unidimensional construct and the distinction between process and content aspects can be a reasonable theoretical and methodological choice (Mullins et al., 2011; Rittle-Johnson et al., 2001; Zimmerman, 2000). Accordingly, models that differentiate between the process and the content aspects of scientific reasoning might have more explanatory value than theories that do not account for the potentially multicomponental nature of scientific reasoning. Accordingly, scientific reasoning should be measured not as a unidimensional construct, rather at least on its process and on its content levels, respectively. On the other hand, further studies should address the question more systematically whether from these two main components (i.e., process and content) an integrated model of scientific reasoning can be developed and to further qualify the relation between those components. At the same time, other (quasi-)experimental studies could investigate if the two aspects of scientific reasoning might interact with each other at different levels of prior (process and content) knowledge or whether the two aspects are rather independent. For example, would low process knowledge and high content knowledge compared to high process knowledge and low content knowledge lead to similar reasoning quality and learning outcomes? In other words, to what extent does high prior knowledge on each of the two aspects of scientific reasoning can compensate for low prior knowledge on the other aspect? Another related question could be whether these two skills should be fostered separately or together?

9.2.2. Conclusions of Study 2

A main finding of the second study was that epistemic network analysis (ENA) was able to detect patterns of scientific reasoning that could meaningfully distinguish between collaborative and individual reasoning processes. Moreover, ENA revealed further information in the data that was hidden from the lenses of the traditional "coding and counting" approach that was applied in Study 1. Namely, ENA was able to identify connections between epistemic processes of scientific reasoning. To be more specific, for dyads evidence evaluation seemed to be the central epistemic process (i.e., making most of the connections with other epistemic processes), and it more often co-occurred with hypothesis generation as well as communicating and scrutinizing than in case of individual reasoning. This outcome suggests that dyads reasoned in a more evidence-based manner than individuals did while considering explanations of the problem and that they more often considered engaging in discourse with others (e.g., colleagues) in order to gain further information and re-evaluate available evidence. These results can indicate, for example, an increased need for sense-making attempts in case of collaborative problem solving, i.e., that reasoning partners engage in a more interactive (Chi, 2009) discussion in order to ask for as well as give clarifications on their hypotheses as well as considering evidence that can support or challenge the different explanations of the problem (Okada & Simon, 1997). These results seem to be in accordance with the assumption that elaborating on the ideas of the reasoning partner or considering engaging others in the reasoning process can also be characteristics of collaborative scientific reasoning (Dunbar, 1995; Okada & Simon, 1997). Interestingly, the lack of connections between communicating and scrutinizing and hypothesis generation on the one hand, and the presence of the already mentioned connection between communication and scrutinizing and evidence evaluation on the other hand, indicate that when dyads took into consideration potential discussions with other reasoners (e.g., they pointed out the relevance of discussing the problem with other teachers), they may not have asked for potential explanations of the problem, but rather for further information in order to solve the problem in a more evidence-based manner (i.e., to find out the performance of the student in other courses). Yet, it is arguable that planned discussions with others (e.g., with colleagues) served the indirect aim to support or reject their initial assumptions and explanations about the problem by finding out evidence that can indirectly confirm or disconfirm their initial assumptions. Further studies might investigate whether such "epistemic chains" or "subnetworks" could be identified that may also reveal further information about how epistemic processes are organized in order to fulfill certain epistemic aims (i.e., to confirm or disconfirm a hypothesis).

Besides, the connections between evidence evaluation and non-epistemic processes might occur, as dyads may have more often made coordinative statements in order to keep up the communication or the problem solving process (Clark & Brennan, 1991), and these statements might have been more often evidence-related (e.g., "Have you read it through?") than e.g., hypothesis-related ("Yes, it can also be that."). It is possible that collaborative reasoners might have made attempts to find what their partner knew, i.e., they coordinated their knowledge (Rummel & Spada, 2005) with respect to the problem case information and for the scientific content information which both served as the available evidence for the study. Further studies should find out whether coordinative attempts in a group may indeed happen more often in the context of some epistemic processes (e.g., evidence evaluation) than in the context of others.

On the other hand, individuals connected evidence evaluation more often with generating solutions than dyads did. This can indicate on the one hand, that individuals were more evidence-based regarding their solutions than dyads were. More specifically, individuals might have engaged more often in conceptual simulation (Trickett & Trafton, 2007) i.e., they might have considered what evidence could indicate the success / failure of their planned solutions (interventions) regarding alleviating the problem. It is also possible that individuals used more often evidence in order to justify their solution plans. On the other hand, however, the connections might also indicate that individuals were more prone to skip an explanatory phase and "jump" to solutions without considering alternative hypotheses or to what extent their initial hypothesis is justifiable. This interpretation might be also supported by the connection between solution generation and hypothesis generation that can indicate that individuals might have coordinated more often their solutions and hypotheses than dyads did because of a lack of initial scrutiny of potential hypotheses about the problem. However, further analyses would be essential in order to confirm the interpretations of the network outcomes. Although ENA provides qualitative options for that purpose, a revised coding scheme could potentially extract sub-codes of e.g., conceptual simulation that, included in further models of ENA, could test between the alternative interpretations of the results of this dissertation.

All in all, the outcomes of the second study suggest that ENA is a potential complementary method to the merely frequency-based "coding and counting" approach. The results of ENA seems not only to reveal further information about the data but also to stimulate more hypotheses for further investigation than an exclusive application of the "coding and counting" approach would lead to.

Moreover, the finding that epistemic networks can reveal interconnected skills of scientific reasoning has important implications regarding how to foster scientific reasoning. Intervention studies, for instance, should take into account that their intervention might not affect merely the epistemic process they target to foster but it might affect a whole network of interconnected skills. For example, fostering evidence evaluation for collaborative reasoners might be beneficial for generating hypotheses but may not be (at least in a direct manner) beneficial for generating solutions. What is more, it is also possible that fostering evidence evaluation would be detrimental for generating solutions without fostering this latter epistemic process at the same time. It is also possible that in case one aims to foster scientific reasoning processes in general, there are certain epistemic processes or a set of processes which might be more effective to target than other processes. For example, those epistemic processes which show more extended connections (e.g., evidence evaluation) might have a larger impact on the overall scientific reasoning than those epistemic processes with limited or no connections with other processes (e.g., drawing conclusions).

To find out how to foster scientific reasoning, one might need, however, some expert model(s) of epistemic networks that should demonstrate how an optimal network should look like. Such expert models can then guide the development of interventions. However, questions would still remain. If, for example, for an expert model hypothesis generation is the central epistemic process while in a group of novices it makes no connections with other epistemic processes, how should one plan the intervention? In this case targeting hypothesis generation may have no effect on other epistemic processes at all. On the other hand, it is also possible that connections would naturally occur and the group of novices would show similar patterns of connections to experts over time. This leads to another implication: namely, that we should understand the development of connections between epistemic processes over time (i.e., the progress of learning to reason scientifically) in order to develop effective and adaptive interventions.

In addition, the predictive power of these networks is a further important question. It is possible that some patterns of connections between epistemic processes can predict better problem solving or other (e.g., domain-specific) learning outcomes from e.g., a problem-based learning or inquiry context. For example, reasoners making more connections between the epistemic processes of questioning, hypothesizing and evaluating evidence while reasoning in a learning environment might demonstrate more elaborated conceptual development compared to reasoners who make fewer connections among these epistemic processes (although they may make more connections among processes). In other words, it is an urgent question to find out to what extent the epistemic network approach of assessing scientific reasoning might contribute to the preparation of future (teacher) professionals to solve problems as scientifically knowledgeable practitioners regarding both aspects, i.e., process and content, of scientific reasoning.

9.3. General Conclusions & Implications for Measurement

The main aim of the present work was to find out to what extent collaboration can facilitate the engagement of future practitioners in scientific reasoning while they are solving authentic problems from their future profession. Investigating this on a sample of pre-service teachers led to mixed results, which allows for the following general implications.

The present work demonstrated that first, scientific reasoning may not be a unidimensional construct and at least two aspects of scientific reasoning should be differentiated: the process and the content aspects. Second, the skills that might contribute to scientific reasoning might be interrelated skills. This brings several implications for the assessment of as well as scaffolding scientific reasoning. For example, both assessment and interventions aiming to measure and foster scientific reasoning should take into account the interrelated nature of skills it might be composed of. Similarly, if one aims to foster one epistemic process (e.g., evidence generation), one should perhaps start to consider that focusing on other epistemic processes (e.g., questioning or hypothesis generation) might lead to the same or even more positive results regarding the development of the epistemic process of interest (in this case, evidence evaluation). Furthermore, taken that scientific reasoning might be composed of a network of interrelated skills, the facilitation of any of those skills might lead to an overspreading activation in the network of interconnected skills. To give an example, this would mean that intervention on hypothesis generation might lead to the activation of the skills which are connected to it, such as evidence evaluation, and even to the activation of further skills which are not "directly" connected to it such as communicating and scrutinizing. Therefore, researchers should consider that their scaffolds might affect the whole network of interrelated skills, and they should design it accordingly instead of focusing on independent epistemic processes. In relation to that, in order to develop scaffolds in order to foster epistemic networks of scientific reasoning, epistemic networks of scientific reasoning should be assessed in order to understand better how the activation of one epistemic process may affect the activation of the whole network of interconnected skills so as to develop interventions accordingly. Yet, it should also be modelled how such activation of the initial networks might predict later development within the network as a result of scaffolding and whether there is an optimal pattern of networks that the scaffolding should target to intervene on. To be more concrete, it might be one possibility that in case of dyads show the problem of connecting evidence with solutions, then scaffolds targeting evidence evaluation and solution generation at the same time would lead to the development of their epistemic networks, i.e., connections between these two epistemic processes would evolve and even connections beyond: leading to more dense networks where, to give an example, drawing conclusions would also become more visible among the interconnected processes.

9.4. Limitations

The main question of the dissertation, whether a collaborative context for problem solving might improve the engagement of future professionals in scientific reasoning, can be answered by the two above reported studies, yet, only to a limited extent. In order to see the explanatory strength of the empirical studies presented in this dissertation (Chapter 8-9) for answering the main question, in the followings, these limitations are addressed.

First, there seems to be an important limitation for generalizability. Although the question of this dissertation is concerned with scientific reasoning of future *practitioners*, the empirical studies investigated scientific reasoning only in the domain or field of *teaching*. It is important to note, that without gaining additional empirical knowledge on how future professionals with other professional background engage in scientific reasoning as a result of collaboration, it is difficult to generalize to other fields of practitioners (e.g., medical professionals). Furthermore, “practitioners” are probably not a homogeneous category: while there might be domain-general characteristics of scientific reasoning, different fields of practice might vary in the extent and strategy they rely on scientific reasoning during solving problems in their field (Ericsson, 2006; Gilhooly, 1990). Yet, it is important to mention here that those domain-specific differences emerge with expertise (Ericsson, 2006) and thus, scientific reasoning of novices from different domains might not show such differences. More empirical research in this area targeting a diversity of domains or fields of practice would contribute to a better understanding on how practitioners reason in different domains, what commonalities they may share and what distinguishes them from each other. By gaining a more evidence-based understanding in this respect, it would be easier to draw conclusions regarding the effect and plan for interventions such as implementing collaboration.

It is noteworthy to add to the generalizability concern that the presented empirical study implemented only one problem case for the problem solving task. In order to estimate the non-task dependent but general scientific reasoning skills of future professionals, future study designs should consider the application of multiple problem solving tasks. In case the results of the present study could be replicated in other problem solving contexts, that could serve as an argument for the external validity (i.e., generalizability) of the present findings.

A further external validity indicator could be if future studies could address more ecologically valid problem solving scenarios. To be more specific, while the research presented in this dissertation was conducted in a laboratory setting, it is important to investigate pre-service teachers problem solving in more authentic (e.g., classroom) settings. By analyzing scientific reasoning of pre-service teachers solving problems stemming from their own teaching experiences would allow us to draw more ecologically valid inferences regarding scientific reasoning and would have the potential of developing more authentic interventions that might require minimal transfer and may result in more effective learning of scientific reasoning skills (e.g., Brown et al., 1989).

A second limitation comes at the interpretation of the results. Most importantly, assessing the quality of scientific reasoning beyond engaging in certain epistemic processes or making reference to scientific content, would be necessary in order to draw concise conclusions regarding the effect of collaboration on scientific reasoning. One potential way to analyze it could be to perhaps combine the process and content aspects and count to what extent the engagement in an epistemic process may contain scientific content reference. As another step, the quality of scientific content reference should be further specified in order to raise its validity as discussed in the conclusions (in section 10.2.1). To give an example, groups' increased engagement in hypothesizing might represent process losses or process gains (Hill, 1982; Hirst & Echterhoff, 2012; Nokes-Malach, et al., 2015), but further analysis should reveal if the increased agreement represents inappropriate explanations as an outcome of a necessary coordination process within groups (process loss) or it is a result of adequate and high-quality argumentation, i.e., more challenges and justifications of ideas (process gain). Although the results of the second study, i.e. the more frequent connections between hypothesizing and evaluating evidence in case of dyads, point to the direction of the latter interpretation (i.e., process gain); without a more explicit operationalization and further measurement on the quality of scientific reasoning it is possible only to a limited extent to draw valid conclusions about whether future professionals reason better in groups than alone.

A third limitation is the lack of a prior knowledge measure that would account for prior content knowledge. Although prior knowledge on the process aspect of scientific reasoning was assessed in order to predict reasoning performance (i.e., engagement in epistemic processes), an analogous measurement of prior *content* knowledge (i.e., about theories and research in the learning sciences) poses limitation for the interpretation of the results regarding the application of scientific content. For example, groups' lower proportional engagement in referring to scientific content can be explained as a lack of knowledge sharing attempts (Stasser & Titus, 1985). Yet, having data on what reasoners knew before they started to reason, would be one way to assess to what extent they shared their knowledge within the group. The importance of this question is visible when one plans an intervention to facilitate scientific reasoning in groups: shall groups receive scaffolds that target knowledge-sharing processes? Further empirical findings are necessary to further warrant the potential implementation of such interventions.

Finally, when explaining the results, it is important to consider that several factors can moderate the effect of collaboration on scientific reasoning. One of those, namely dyadic heterogeneity regarding problem solving scripts, was explicitly addressed in this work. Yet, studies show that further factors such as collaboration skills (i.e., internal collaboration scripts; Fischer et al., 2013) as well as cognitive load (e.g., Kirschner et al., 2011) may account for the extent to which collaborative partners engage in productive discussion. These moderators may also influence the extent to which dyadic heterogeneity regarding problem solving scripts can explain scientific reasoning. For example, heterogeneous dyads might face more coordination demands to develop a mutual understanding on the problem, and this might impose more cognitive load on dyadic members which might not allow them to benefit from a collaborative discourse. On the other hand, coordination demands stemming from dyadic heterogeneity might be minimized in the presence of strong collaborative skills. Such interaction or moderation effects should be enlightened by further empirical investigations.

9.5. Practical Implications

The findings of this dissertation indicated that implementing collaboration can be beneficial if the aim is to engage future professionals in a more reflective and scrutinized understanding of the nature of the problem (i.e., developing hypotheses and connecting those with evidence). At the same time, it might be necessary to scaffold future practitioners to apply scientific evidence from their domain, because engaging them in collaboration with minimal instruction does not have a positive effect on applying scientific content. These scaffolds could potentially target sharing unique knowledge with each other while solving the problem as well as regulating coordination and sense-making processes at an early phase of problem solving within the group.

Furthermore, it seems that this reflective understanding of the nature of the problem can further benefit from group heterogeneity. Therefore, when planning for a collaborating problem solving session (e.g., initiating problem based learning) it seems to be a reasonable idea to pair collaborating partners who to an extent disagree with each other on how to progress with the problem solving itself. Such disagreements, according to the present empirical work, seemed to stimulate a more investigative discussion about the problem. However, another implication of the presented findings is that disagreeing partners may require further scaffolding to regulate coordination processes as well.

On the other hand, if the aim is to support practitioners to reason about how to solve the problem instead of explaining it, individual reasoning might be more appropriate than reasoning in groups. Yet, individuals might need to be supported to re-evaluate their initial interpretations of the problem by considering alternative explanations for it in the light of evidence.

All in all, collaboration might be beneficial for engaging future professionals in reasoning as scientifically knowledgeable practitioners, but further instructional support for sharing ideas, coordinating views and initiating solutions seem to be necessary to optimize collaborative problem solving. The type and level of instructional support should be further qualified though.

References

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior, 22*(3), 261–295. https://doi.org/10.1016/S0022-5371(83)90201-3

Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem situations. *Psychological Review*, *94*(2), 192–210. https://doi.org/10.1037//0033-295X.94.2.192

Anderson, J. R., & Fincham, J. M. (2014). Discovering the Sequential Structure of Thought. *Cognitive Science*, *38*(2), 322–352. https://doi.org/10.1111/cogs.12068

Arastoopour, G., Shaffer, D. W., Swiecki, Z., Ruis, A. R., & Chesler, N. C. (2016). Teaching and assessing engineering design thinking with virtual internships and epistemic network analysis. *International Journal of Engineering Education, 32*(3), 1492–1501.

Archer, L., DeWitt, J., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2010). “Doing” science versus “being” a scientist: Examining 10/11-year-old schoolchildren’s constructions of science through the lens of identity. *Science Education*, *94*(4), 617–639. https://doi.org/10.1002/sce.20399

Asim, A., Ekuri, E., & Eni, E. (2013). A diagnostic study of pre-service teachers’ competency in multiple-choice item development. *Research in Education*, *89*(1), 13–22. https://doi.org/10.7227/RIE.89.1.2

Asterhan, C. S. C., & Schwarz, B. B. (2009). Argumentation and Explanation in Conceptual Change: Indications From Protocol Analyses of Peer-to-Peer Dialog. *Cognitive Science*, *33*(3), 374–400. https://doi.org/10.1111/j.1551-6709.2009.01017.x

Aviv, R., Erlich, Z., Ravid, G., & Geva, A. (2003). Network analysis of knowledge construction in asynchronous learning networks. *Journal of Asynchronous Learning Networks*, *7*(3), 1–23.

Baeten, M., & Simons, M. (2014). Student teachers' team teaching: Models, effects, and conditions for implementation. *Teaching and Teacher Education*, *41*, 92–110. https://doi.org/10.1016/j.tate.2014.03.010

Baeten, M., & Simons, M. (2016). Student teachers' team teaching: how do learners in the classroom experience team-taught lessons by student teachers? *Journal of Education for Teaching*, *42*(1), 93–105. https://doi.org/10.1080/02607476.2015.1135226

Bagley, E., & Shaffer, D.W. (2015). Learning in an urban and regional planning practicum: The view from educational ethnography. *Journal of Interactive Learning Research, 26(4),* 369-393.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511527685

Berkowitz, M.W., & Gibbs, J.C. (1983). Measuring the developmental features of moral discussion. *Merrill-Palmer Quarterly, 29*, 399-410.

Barron, B. (2000). Achieving Coordination in Collaborative Problem-Solving Groups. *Journal of the Learning Sciences*, *9*(4), 403–436. https://doi.org/10.1207/S15327809JLS0904_2

Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift Für Erziehungswissenschaft*, *9*(4), 469–520.

Birrell, J. R., & Bullough, R. V. (2005). Teaching with a Peer: A Follow-Up Study of the 1st Year of Teaching. *Action in Teacher Education*, *27*(1), 72–81. https://doi.org/10.1080/01626620.2005.10463375

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network Analysis in the Social Sciences. *Science*, *323*(5916), 892–895. https://doi.org/10.1126/science.1165821

Bowers, C. A., Pharmer, J. A., & Salas, E. (2000). When member homogeneity is needed in work teams: A meta-analysis. *Small Group Research*, *31*(3), 305–327. https://doi.org/10.1177/104649640003100303

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated Cognition and the Culture of Learning. *Educational Researcher*, *18*(1), 32–42. https://doi.org/10.3102/0013189X018001032

Brown, N. J. S., Furtak, E. M., Timms, M., Nagashima, S. O., & Wilson, M. (2010). The Evidence-Based Reasoning Framework: Assessing Scientific Reasoning. *Educational Assessment, 15*(3–4), 123–141. https://doi.org/10.1080/10627197.2010.530551

Bullough, R. V., Young, J., Erickson, L., Birrell, J. R., Clark, D. C., Egan, M. W., … Smith, G. (2002). Rethinking Field Experience: Partnership Teaching Versus Single-Placement Teaching. *Journal of Teacher Education*, *53*(1), 68–80. https://doi.org/10.1177/0022487102053001007

Canham, M. S., Wiley, J., & Mayer, R. E. (2012). When diversity in training improves dyadic problem solving. *Applied Cognitive Psychology*, *26*(3), 421–430. https://doi.org/10.1002/acp.1844

Charlin, B., Boshuizen, H. P. A., Custers, E. J., & Feltovich, P. J. (2007). Scripts and clinical reasoning: clinical expertise. *Medical Education*, *41*(12), 1178–1184. https://doi.org/10.1111/j.1365-2923.2007.02924.x

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition.* (pp. 127–

149). Washington: American Psychological Association. https://doi.org/10.1037/10096-006

Csanadi, A., Daxenberger, J., Ghanem, C., Kollar, I., Fischer, F., & Gurevych, I. (2016, July). Automated Text Classification to Capture Scientific Reasoning and Argumentation Processes in Different Professional Problem Solving Contexts. Paper presented at the Twenty-sixth Annual Meeting of the Society for Text & Discourse, Kassel, Germany.

Csanadi, A., Kollar, I., Fischer, F. (2016). Scientific reasoning and problem solving in a practical domain: Are two heads better than one? In C.-K. Looi, J. Polman, U. Cress, & P. Reimann (Eds.), *Transforming Learning, Empowering Learners: The International Conference of the Learning Sciences (ICLS) 2016: Volume 1* (pp 50-57). Singapore: International Society of the Learning Sciences.

Charlin, B., Boshuizen, H. P. A., Custers, E. J., & Feltovich, P. J. (2007). Scripts and clinical reasoning: clinical expertise. *Medical Education*, *41*(12), 1178–1184. https://doi.org/10.1111/j.1365-2923.2007.02924.x

Chi, M. T., & Koeske, R. D. (1983). Network representation of a child's dinosaur knowledge. *Developmental Psychology*, *19*(1), 29–39. https://doi.org/10.1037/0012-1649.19.1.29

Chi, Michelene T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *Journal of the Learning Sciences*, *6*(3), 271–315. https://doi.org/10.1207/s15327809jls0603_1

Chi, Michelene T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*(1), 73–105. https://doi.org/10.1111/j.1756-8765.2008.01005.x

Chi, Michelene T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and Representation of Physics Problems by Experts and Novices. *Cognitive Science*, *5*(2), 121–152. https://doi.org/10.1207/s15516709cog0502_2

Chi, Michelene T. H., Siler, S. A., & Jeong, H. (2004). Can Tutors Monitor Students' Understanding Accurately? *Cognition and Instruction*, *22*(3), 363–387. https://doi.org/10.1207/s1532690xci2203_4

Chi, Michelene T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Chi, M.T.H. (2008). Three Types of Conceptual Change: Belief Revision, Mental Model Transformation, and Categorical Shift. In S. Vosniadou, *International Handbook of Research on Conceptual Change* (pp. 61–82). New York: Routledge.

Chinn, C. A., & Brewer, W. F. (1993). The Role of Anomalous Data in Knowledge Acquisition: A Theoretical Framework and Implications for Science Instruction. *Review of Educational Research*, *63*(1), 1–49. https://doi.org/10.2307/1170558

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, *86*(2), 175–218. https://doi.org/10.1002/sce.10001

Chiu, M. M. (2005). A New Method for Analyzing Sequential Processes: Dynamic Multilevel Analysis. *Small Group Research*, *36*(5), 600–631. https://doi.org/10.1177/1046496405279309

Chiu, M. M. (2008). Flowing toward Correct Contributions during Group Problem Solving: A Statistical Discourse Analysis. *The Journal of the Learning Sciences*, *17*(3), 415–463.

Clarà, M., & Mauri, T. (2010). Toward a dialectic relation between the results in CSCL: Three critical methodological aspects of content analysis schemes. *International Journal of Computer-Supported Collaborative Learning,* 5(1), 117-136.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition.* (pp. 127–149). Washington: American Psychological Association. https://doi.org/10.1037/10096-006

Clark, H. H., & Schaefer, E. F. (1989). Contributing to Discourse. *Cognitive Science*, *13*(2), 259–294. https://doi.org/10.1207/s15516709cog1302_7

Collier, W., Ruis, A. R., & Shaffer, D. W. (2016). Local versus global connection making in discourse. In C. K. Looi, J. L. Polman, U. Cress, & P. Reimann (Eds.), Transforming Learning, Empowering Learners: *The International Conference of the Learning Sciences (ICLS) 2016: Volume 1* (pp 426-433). Singapore: International Society of the Learning Sciences.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 240–247. https://doi.org/10.1016/S0022-5371(69)80069-1

Cress, U., & Hesse, W. (2013). Quantitative methods for studying small groups. In C. A. Hmelo-Silver, C. Chinn, C. Chan, & A. M. O'Donnell, *The international handbook of collaborative learning* (pp. 93–111). New York: Routledge.

Curşeu, P. L., & Pluut, H. (2013). Student groups as learning entities: The effect of group diversity and teamwork quality on groups' cognitive complexity. *Studies in Higher Education*, *38*(1), 87–103. https://doi.org/10.1080/03075079.2011.565122

de Jong, T. (2006). Computer Simulations: Technological Advances in Inquiry Learning. *Science*, *312* (5773), 532–533. https://doi.org/10.1126/science.1127750

de Jong, T., & Van Joolingen, W. R. (1998). Scientific Discovery Learning with Computer Simulations of Conceptual Domains. *Review of Educational Research*, *68*(2), 179–201. https://doi.org/10.3102/00346543068002179

de Laat, M., Lally, V., Lipponen, L., & Simons, R.-J. (2007). Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social Network Analysis. *International Journal of Computer-Supported Collaborative Learning*, *2*(1), 87–103. https://doi.org/10.1007/s11412-007-9006-4

Dennen, V. P., & Hoadley, C. (2013). Designing collaborative learning through computer-support. In C. Hmelo-Silver, C. Chinn, C. K. K. Chan & A. M. O'Donnel (Eds.). *The International Handbook of Collaborative Learning* (pp. 389-402). New York: Routledge

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology*, *53*(3), 497–509. https://doi.org/10.1037/0022-3514.53.3.497

Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg, J. E. Davidson, R. J. Sternberg (Ed), & J. E. Davidson (Eds.), *The nature of insight* (pp. 365–395). Cambridge, MA: The MIT Press.

Eells, T. D., Lombart, K. G., Salsman, N., Kendjelic, E. M., Schneiderman, C. T., & Lucas, C. P. (2011). Expert reasoning in psychotherapy case formulation. *Psychotherapy Research*, *21*(4), 385–399. https://doi.org/10.1080/10503307.2010.539284

Elbers, E., & Streefland, L. (2000). Collaborative learning and the construction of common knowledge. *European Journal of Psychology of Education*, *15*(4), 479–490. https://doi.org/10.1007/BF03172989

Engelmann, K., Neuhaus, B. J., & Fischer, F. (2016). Fostering scientific reasoning in education–meta-analytic evidence from intervention studies. *Educational Research and Evaluation*, *22*(5-6), 333-349.

Ericsson, K. A. (2006). Protocol Analysis and Expert Thought: Concurrent Verbalizations of Thinking during Experts' Performance on Representative Tasks. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 223–242). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511816796.013

Erkens, G., Kanselaar, G., Prangsma, M., & Jaspers, J. (2003). Computer support for collaborative and argumentative writing. In E. de Corte, L. Verschaffel, N. Entwistle, & J. Van Merriënboer (Eds.), *Advances in learning and instruction series. Powerful learning environments: Unravelling basic components and dimensions* (pp 159-177). Oxford, England: Pergamon/Elsevier Science Ltd.

Fischer, F., Kollar, I., Stegmann, K., & Wecker, C. (2013). Toward a script theory of guidance in computer-supported collaborative learning. *Educational Psychologist*, *48*(1), 56–66. https://doi.org/10.1080/00461520.2012.748005

Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., Neuhaus, B., Dorner, B., Pankofer, S., Fischer, M., Strijbos, J-W., Heene, M., Eberle, J. (2014). Scientific

Reasoning and Argumentation: Advancing an Interdisciplinary Research Agenda in Education. *Frontline Learning Research*, *2*(3), 28–45. https://doi.org/10.14786/flr.v2i2.96

Fischer, F., & Mandl, H. (2005). Knowledge Convergence in Computer-Supported Collaborative Learning: The Role of External Representation Tools. *Journal of the Learning Sciences*, *14*(3), 405–441. https://doi.org/10.1207/s15327809jls1403_3

Gardiner, W., & Robinson, K. S. (2009). Paired Field Placements: A Means for Collaboration. *New Educator, 5*(1), 81–94.

Gee, J. P. (1999). *An introduction to discourse analysis.* New York: Routledge.

Gilhooly, K. J. (1990). Cognitive psychology and medical diagnosis. *Applied Cognitive Psychology*, *4*(4), 261–272. https://doi.org/10.1002/acp.2350040404

Goldman, S. R., Braasch, J. L. G., Wiley, J., Graesser, A. C., & Brodowinska, K. (2012). Comprehending and Learning from Internet Sources: Processing Patterns of Better and Poorer Learners. *Reading Research Quarterly*, *47*(4), 356–381.

Gordon, D. M., Collins, J. C., & Jewell, K. L. (2015). Project-based Approach to Teaching Standardized Test Construction: A Model for Pre-Service Education Students on How to Standardize a Test. *American Journal of Educational Research, American Journal of Educational Research*, *3*(12), 1528–1535. https://doi.org/10.12691/education-3-12-8

Gräsel, C., & Mandl, H. (1993). Förderung des Erwerbs diagnostischer Strategien in fallbasierten Lernumgebungen. *Unterrichtswissenschaft*, *21*(4), 355-369.

Grissom, R. J. (2000). Heterogeneity of variance in clinical data. *Journal of Consulting and Clinical Psychology*, *68*(1), 155–165. https://doi.org/10.1037/0022-006X.68.1.155

Gruber, H., Mandl, H., & Renkl, A. (2000). Was lernen wir in Schule und Hochschule: Träges Wissen? In H. Mandl & J. Gerstenmeier (Eds.), *Die Kluft zwischen Wissen und Handeln : empirische und theoretische Lösungsansätze* (pp. 139–156). Göttingen: Hogrefe.

Hatfield, D. (2015). The Right Kind Of Telling:: An Analysis of Feedback and Learning in a Journalism Epistemic Game. *International Journal of Gaming and Computer-Mediated Simulations, 7*(2), 1–23. https://doi.org/10.4018/IJGCMS.2015040101

Hausmann, R. G. M., Chi, M. T. H., & Roy, M. (2004). Learning from collaborative problem solving: An analysis of three hypothesized Mechanisms. *Proceedings of the Cognitive Science Society*, *26*(26), 547-552.

Hetmanek, A., Wecker, C., Kiesewetter, J., Trempler, K., Fischer, M. R., Gräsel, C., & Fischer, F. (2015). Wozu nutzen Lehrkräfte welche Ressourcen? Eine Interviewstudie zur Schnittstelle zwischen bildungswissenschaftlicher Forschung und professionellem Handeln im Bildungsbereich. *Unterrichtswissenschaft*, *43*(3), 193–208.

Hill, G. W. (1982). Group versus individual performance: Are N+1 heads better than one? *Psychological Bulletin*, *91*(3), 517–539. https://doi.org/10.1037/0033-2909.91.3.517

Hinsley, D. A., Hayes, J. R., & Simon, H. A. (1977). From words to equations: Meaning and representation in algebra word problems. In P. A. Carpenter & M. A. Just (Eds.), *Cognitive processes in comprehension* (p. 89-106). New Jersey: Lawrence Erlbaum Associates

Hinsz, V. B., Tindale, R. S., & Vollrath, D. A. (1997). The emerging conceptualization of groups as information processors. *Psychological Bulletin*, *121*(1), 43–64.

Hirst, W., & Echterhoff, G. (2012). Remembering in Conversations: The Social Sharing and Reshaping of Memories. *Annual Review of Psychology*, *63*(1), 55–79. https://doi.org/10.1146/annurev-psych-120710-100340

Hmelo-Silver, C. E. (2004). Problem-Based Learning: What and How Do Students Learn? *Educational Psychology Review*, *16*(3), 235–266. https://doi.org/10.1023/B:EDPR.0000034022.16470.f3

Hmelo-Silver, C. E., Marathe, S., & Liu, L. (2007). Fish Swim, Rocks Sit, and Lungs Breathe: Expert-Novice Understanding of Complex Systems. *Journal of the Learning Sciences*, *16*(3), 307–331. https://doi.org/10.1080/10508400701413401

Hollingshead, A. B., & Brandon, D. P. (2003). Potential Benefits of Communication in Transactive Memory Systems. *Human Communication Research*, *29*(4), 607–615. https://doi.org/10.1111/j.1468-2958.2003.tb00859.x

Horton, W. S., & Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition*, *96*(2), 127–142. https://doi.org/10.1016/j.cognition.2004.07.001

Hou, H.-T., Sung, Y.-T., & Chang, K.-E. (2009). Exploring the behavioral patterns of an online knowledge-sharing discussion activity among teachers with problem-solving strategy. *Teaching and Teacher Education*, *25*(1), 101–108. https://doi.org/10.1016/j.tate.2008.07.006

Hutchins, E. (1995). How a Cockpit Remembers Its Speeds. *Cognitive Science*, *19*(3), 265–288. https://doi.org/10.1207/s15516709cog1903_1

Jang, S.-J. (2008). Innovations in science teacher education: Effects of integrating technology and team-teaching strategies. *Computers & Education*, *51*(2), 646–659. https://doi.org/10.1016/j.compedu.2007.07.001

Jeong, A. (2005). A Guide to Analyzing Message–Response Sequences and Group Interaction Patterns in Computer-mediated Communication. *Distance Education*, *26*(3), 367–383. https://doi.org/10.1080/01587910500291470

Jeong, A. C. (2003). The Sequential Analysis of Group Interaction and Critical Thinking in Online. *American Journal of Distance Education*, *17*(1), 25–43. https://doi.org/10.1207/S15389286AJDE1701_3

Kamens, M. W. (2007). Learning about Co-Teaching: A Collaborative Student Teaching Experience for Preservice Teachers. *Teacher Education and Special Education*, *30*(3).

Kirschner, F., Paas, F., & Kirschner, P. A. (2009). A Cognitive Load Approach to Collaborative Learning: United Brains for Complex Tasks. *Educational Psychology Review*, *21*(1), 31–42. https://doi.org/10.1007/s10648-008-9095-2

Kirschner, F., Paas, F., Kirschner, P. A., & Janssen, J. (2011). Differential effects of problem-solving demands on individual and collaborative learning outcomes. *Learning and Instruction*, *21*(4), 587–599. https://doi.org/10.1016/j.learninstruc.2011.01.001

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, *41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Klahr, D., Zimmerman, C., & Jirout, J. (2011). Educational Interventions to Advance Children's Scientific Thinking. *Science*, *333*(6045), 971–975. https://doi.org/10.1126/science.1204528

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, *12*(1), 1–48. https://doi.org/10.1207/s15516709cog1201_1

Klahr, D., & Dunbar, K. (1989). Developmental differences in scientific discovery processes. In *Complex information processing: The impact of Herbert A. Simon. Hillsdale* (pp. 109–143). NJ : Lawrence Erlbaum Associates.

Klahr, D., & Simon, H. A. (1999). Studies of scientific discovery: Complementary approaches and convergent findings. *Psychological Bulletin*, *125*(5), 524–543. https://doi.org/10.1037/0033-2909.125.5.524

Klopp, E., Stark, R., Kopp, V., & Fischer, M. R. (2013). Psychological Factors Affecting Medical Students' Learning with Erroneous Worked Examples. *Journal of Education and Learning*, *2*(1), 158–170.

Koenig, K., Schen, M., & Bao, L. (2012). Explicitly Targeting Pre-Service Teacher Scientific Reasoning Abilities and Understanding of Nature of Science through an Introductory Science Course. *Science Educator*, *21*(2), 1.

König, J., Blömeke, S., Paine, L., Schmidt, W. H., & Hsieh, F. J. (2011). General pedagogical knowledge of future middle school teachers: On the complex ecology of teacher education in the United States, Germany, and Taiwan. *Journal of Teacher Education*, *62*(2), 188–201. https://doi.org/10.1177/0022487110388664

Kollar, I., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction*, *17*(6), 708–721. https://doi.org/10.1016/j.learninstruc.2007.09.021

Kruger, A. C., & Tomasello, M. (1986). Transactive discussions with peers and adults. *Developmental Psychology*, *22*(5), 681–685. https://doi.org/10.1037/0012-1649.22.5.681

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, *96*(4), 674–689.

Kuhn, D., & Dean, D. (2005). Is Developing Scientific Thinking All About Learning to Control Variables? *Psychological Science*, *16*(11), 866–870. https://doi.org/10.1111/j.1467-9280.2005.01628.x

Kuhn, D. (2001). How do People Know? *Psychological Science*, *12*(1), 1–8. https://doi.org/10.1111/1467-9280.00302

Kuhn, D. (2002). What is Scientific Thinking and How Does It Develop? In U. Goswami (Ed.), *Blackwell Handbook of Childhood Cognitive Development* (pp. 371–393). Malden, MA, USA: Blackwell Publishers Ltd. https://doi.org/10.1002/9780470996652.ch17

Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511815355

Lazonder, A. W. (2005). Do two heads search better than one? Effects of student collaboration on web search behaviour and search outcomes. *British Journal of Educational Technology*, *36*(3), 465–475. https://doi.org/10.1111/j.1467-8535.2005.00478.x

Lawson, A. E. (2002). Sound and faulty arguments generated by preservice biology teachers when testing hypotheses involving unobservable entities. *Journal of Research in Science Teaching*, *39*(3), 237–252. https://doi.org/10.1002/tea.10019

Lockhorst, D., Wubbels, T., & Oers, B. van. (2010). Educational dialogues and the fostering of pupils' independence: the practices of two teachers. *Journal of Curriculum Studies*, *42*(1), 99–121. https://doi.org/10.1080/00220270903079237

Madsen, A. J., & Olson, J. K. (2005). Student teachers' use of learning theories to diagnose children's learning difficulties. *Journal of Elementary Science Education*, *17*(1), 55. https://doi.org/10.1007/BF03174673

Mandinach, E. B. (2012). A Perfect Time for Data Use: Using Data-Driven Decision Making to Inform Practice. *Educational Psychologist*, *47*(2), 71–85. https://doi.org/10.1080/00461520.2012.667064

McElvany, N., Schroeder, S., Baumert, J., Schnotz, W., Horz, H., & Ullrich, M. (2012). Cognitively demanding learning materials with texts and instructional pictures: teachers' diagnostic skills, pedagogical beliefs and motivation. *European Journal of Psychology of Education*, *27*(3), 403–420. https://doi.org/10.1007/s10212-011-0078-1

McNeill, K. L., & Knight, A. M. (2013). Teachers' pedagogical content knowledge of scientific argumentation: The impact of professional development on K–12 teachers. *Science Education*, *97*(6), 936–972. https://doi.org/10.1002/sce.21081

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*(02), 57–74. https://doi.org/10.1017/S0140525X10000968

Métrailler, Y. A., Reijnen, E., Kneser, C., & Opwis, K. (2008). Scientific problem solving in a virtual laboratory: A comparison between individuals and pairs. *Swiss Journal of Psychology / Schweizerische Zeitschrift Für Psychologie / Revue Suisse de Psychologie*, *67*(2), 71–83. https://doi.org/10.1024/1421-0185.67.2.71

Meyers, L. S., Gamst, G., & Guarino, A. J. (2006). *Applied Multivariate Research: Design and Interpretation*. SAGE.

Miura, A., & Hida, M. (2004). Synergy between diversity and similarity in group-idea generation. *Small Group Research*, *35*(5), 540–564. https://doi.org/10.1177/1046496404264942

Morris, A. K. (2006). Assessing pre-service teachers' skills for analyzing teaching. *Journal of Mathematics Teacher Education*, *9*(5), 471–505. https://doi.org/10.1007/s10857-006-9015-7

Morrison, J. A., & Lederman, N. G. (2003). Science teachers' diagnosis and understanding of students' preconceptions. *Science Education*, *87*(6), 849–867. https://doi.org/10.1002/sce.10092

Mullins, D., Rummel, N., & Spada, H. (2011). Are two heads always better than one? Differential effects of collaboration on students' computer-supported learning in mathematics. *International Journal of Computer-Supported Collaborative Learning*, *6*(3), 421–443. https://doi.org/10.1007/s11412-011-9122-z

Nevin, A. I., Thousand, J. S., & Villa, R. A. (2009). Collaborative teaching for teacher educators – What does the research say? *Teaching and Teacher Education, 25*(4), 569–574. https://doi.org/10.1016/j.tate.2009.02.009

Nokes, J. D., Bullough Jr., R. V., Egan, W. M., Birrell, J. R., & Merrell Hansen, J. (2008). The paired-placement of student teachers: An alternative to traditional placements in secondary schools. *Teaching and Teacher Education*, *24*(8), 2168–2177. https://doi.org/10.1016/j.tate.2008.05.001

Nokes-Malach, T. J., Richey, J. E., & Gadgil, S. (2015). When Is It Better to Learn Together? Insights from Research on Collaborative Learning. *Educational Psychology Review*, *27*(4), 645–656. https://doi.org/10.1007/s10648-015-9312-8

Noroozi, O., Teasley, S. D., Biemans, H. J. A., Weinberger, A., & Mulder, M. (2013). Facilitating learning in multidisciplinary groups with transactive CSCL scripts. *International Journal of Computer-Supported Collaborative Learning*, *8*(2), 189–223. https://doi.org/10.1007/s11412-012-9162-z

Okada, T., & Simon, H. A. (1997). Collaborative discovery in a scientific domain. *Cognitive Science*, *21*(2), 109–146. https://doi.org/10.1016/S0364-0213(99)80020-2

Osborne, J. (2010). Arguing to Learn in Science: The Role of Collaborative, Critical Discourse. *Science*, *328*(5977), 463–466. https://doi.org/10.1126/science.1183944

Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical assessment, research & evaluation*, *9*(6), 1-12.

Oshima, J., Oshima, R., & Matsuzawa, Y. (2012). Knowledge Building Discourse Explorer: a social network analysis application for knowledge building discourse. *Educational Technology Research and Development*, *60*(5), 903–921. https://doi.org/10.1007/s11423-012-9265-2

Palonen, T., & Hakkarainen, K. (n.d.). Patterns of Interaction in Computer-Supported Learning: A Social Network Analysis. In B. Fishman & S. O'Connor-Divelbiss (Eds.), *Fourth International Conference of the Learning Sciences* (pp. 334–339). NJ: Erlbaum: Mahwah.

Paulus, P. (2000). Groups, Teams, and Creativity: The Creative Potential of Idea-generating Groups. *Applied Psychology*, *49*(2), 237–262. https://doi.org/10.1111/1464-0597.00013

Pavitt, C., & Johnson, K. K. (2001). The association between group procedural MOPs and group discussion procedure. *Small Group Research*, *32*(5), 595–624. https://doi.org/10.1177/104649640103200505

Pea, R. D. (1993). Practices of distributed intelligences and design for education. In G. Solomon (Ed*.), Distributed cognitions: Psychological and educational considerations* (pp. 47-87). Cambridge, UK: Cambridge University Press.

Pickthorne, B. (1983). Error factors: A missing link between cognitive science and classroom practice? *Instructional Science*, *11*(4), 281–312. https://doi.org/10.1007/BF00137291

Robinson, V. M. J. (1998). Methodology and the Research-Practice Gap. *Educational Researcher*, *27*(1), 17–26. https://doi.org/10.3102/0013189X027001017

Pol, J. van de, Volman, M., Oort, F., & Beishuizen, J. (2014). Teacher Scaffolding in Small-Group Work: An Intervention Study. *Journal of the Learning Sciences*, *23*(4), 600–650. https://doi.org/10.1080/10508406.2013.805300

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology*, *93*(2), 346–362. https://doi.org/10.1037/0022-0663.93.2.346

Roschelle, J., & Teasley, S. D. (1995). The Construction of Shared Knowledge in Collaborative Problem Solving. In C. O'Malley (Ed.), *Computer Supported Collaborative Learning* (pp. 69–97). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-85098-1_5

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W., (2000). Methodological Issues in the Content Analysis of Computer Conference Transcripts. *International Journal of Artificial Intelligence in Education, 12*, 8-22.

Rummel, N., & Spada, H. (2005). Learning to Collaborate: An Instructional Approach to Promoting Collaborative Problem Solving in Computer-Mediated Settings. *Journal of the Learning Sciences*, *14*(2), 201–241. https://doi.org/10.1207/s15327809jls1402_2

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, *25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020

Sackett, D. L., Rosenberg, W. M., Gray, J. A., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. *BMJ : British Medical Journal*, *312*(7023), 71–72. https://doi.org/10.1136/bmj.312.7023.71

Sampson, V., & Blanchard, M. R. (2012). Science teachers and scientific argumentation: Trends in views and practice. *Journal of Research in Science Teaching*, *49*(9), 1122–1148. https://doi.org/10.1002/tea.21037

Sampson, V., & Clark, D. (2009). The impact of collaboration on the outcomes of scientific argumentation. *Science Education*, *93*(3), 448–484. https://doi.org/10.1002/sce.20306

Sampson, V., & Clark, D. B. (2008). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education*, *92*(3), 447–472. https://doi.org/10.1002/sce.20276

Sandoval, W. A. (2005). Understanding students' practical epistemologies and their influence on learning through inquiry. *Science Education*, *89*(4), 634–656. https://doi.org/10.1002/sce.20065

Sandoval, W. A., & Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, *88*(3), 345–372. https://doi.org/10.1002/sce.10130

Sandoval, W. A., Sodian, B., Koerber, S., & Wong, J. (2014). Developing Children's Early Competencies to Engage With Science. *Educational Psychologist*, *49*(2), 139–152. https://doi.org/10.1080/00461520.2014.917589

Satterfield, J. M., Spring, B., Brownson, R. C., Mullen, E. J., Newhouse, R. P., Walker, B. B., & Whitlock, E. P. (2009). Toward a Transdisciplinary Model of Evidence-Based Practice: *A Transdisciplinary Model of Evidence-Based Practice*. *Milbank Quarterly*, *87*(2), 368–390. https://doi.org/10.1111/j.1468-0009.2009.00561.x

Schank, R. C. (1999). *Dynamic memory revisited*. New York, NY, US: Cambridge University Press. https://doi.org/10.1017/CBO9780511527920

Schauble, L., Glaser, R., Raghavan, K., & Reiner, M. (1991). Causal Models and Experimentation Strategies in Scientific Reasoning. *Journal of the Learning Sciences*, *1*(2), 201–238. https://doi.org/10.1207/s15327809jls0102_3

Schunn, C. D., & Anderson, J. R. (1999). The Generality/Specificity of Expertise in Scientific Reasoning. *Cognitive Science*, *23*(3), 337–370. https://doi.org/10.1207/s15516709cog2303_3

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A Tutorial on Epistemic Network Analysis: Analyzing the Structure of Connections in Cognitive, Social, and Interaction Data. *Journal of Learning Analytics*, *3*(3), 9–45.

Shaffer, D. W., Hatfield, D. L., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E. A., … Frank, K. (2009). Epistemic Network Analysis: A prototype for 21st century assessment of learning. *International Journal of Learning and Media*, *1*(1), 1–21.

Shapiro, A. M. (2004). How Including Prior Knowledge As a Subject Variable May Change Outcomes of Learning Research. *American Educational Research Journal*, *41*(1), 159–189. https://doi.org/10.3102/00028312041001159

Shehab, H. M., & Nussbaum, E. M. (2015). Cognitive load of critical thinking strategies. *Learning and Instruction*, *35*, 51–61. https://doi.org/10.1016/j.learninstruc.2014.09.004

Shin, M., Lee, H., & McKenna, J. W. (2016). Special education and general education preservice teachers' co-teaching experiences: a comparative synthesis of qualitative research. *International Journal of Inclusive Education*, *20*(1), 91–107. https://doi.org/10.1080/13603116.2015.1074732

Siebert-Evenstone, A. L., Arastoopour, G., Collier, W., Swiecki, W., Ruis, A. R., & Shaffer, D. W. (2016). In search of conversational grain size: Modeling semantic structure using moving stanza windows. In C. K. Looi, J. L. Polman, U. Cress, & P. Reimann (Eds.), *Transforming Learning, Empowering Learners: The International Conference of the Learning Sciences (ICLS) 2016: Volume 1* (pp 631–638). Singapore: International Society of the Learning Sciences.

Siegler, R. S. (1978). The origins of scientific reasoning. In R. S. Siegler & R. S. Siegler (Eds.), *Children's thinking: What develops?* (pp. 109–149). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Simon HA. 1989. The scientist as problem solver. In D. Klahr, K. Kotovsky (Eds.) Complex Information Processing: The Impact of Herbert A. Simon (pp. 373–98). Hillsdale, NJ: Erlbaum

Simon, H. A., Klahr, D., & Kotovsky, K. (Eds.). (1989). *Complex information processing: the impact of Herbert A. Simon*. Hillsdale, N.J: L. Erlbaum Associates.

Simon, H. A., Langley, P. W., & Bradshaw, G. L. (1981). Scientific Discovery as Problem Solving. *Synthese*, *47*(1), 3–14.

Simon, H. A., & Lea, G. (1974). Problem solving and rule induction: A unified view. In L. W. Gregg (Ed.), *Knowledge and cognition* (pp 105-128). Oxford, England: Lawrence Erlbaum.

Spencer, T. D., Detrich, R., & Slocum, T. A. (2012). Evidence-based Practice: A Framework for Making Effective Decisions. *Education and Treatment of Children*, *35*(2), 127–151. https://doi.org/10.1353/etc.2012.0013

Stark, R., Gruber, H., Mandl, H., & Hinkofer, L. (2001). Wege zur Optimierung eines beispielbasierten Instruktionsansatzes: Der Einfluss multipler Perspektiven und instruktionaler Erklärungen auf den Erwerb von Handlungskompetenz. *Unterrichtswissenschaft*, *29*(1), 26–40.

Stark, R., Gruber, H., Mandl, H., & Renkl, A. (1998). Indeed, sometimes knowledge does not help: A replication study. *Instructional Science*, *26*(5), 391–407. https://doi.org/10.1023/A:1003209514232

Stark, R., Puhl, T., & Krause, U.-M. (2009). Improving scientific argumentation skills by a problem-based learning environment: effects of an elaboration tool and relevance of student characteristics. *Evaluation & Research in Education*, *22*(1), 51–68. https://doi.org/10.1080/09500790903082362

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology*, *48*(6), 1467–1478. https://doi.org/10.1037/0022-3514.48.6.1467

Strijbos, J.-W., Martens, R. L., Jochems, W. M. G., & Broers, N. J. (2004). The Effect of Functional Roles on Group Efficiency: Using Multilevel Modeling and Content Analysis to Investigate Computer-Supported Collaboration in Small Groups. *Small Group Research*, *35*(2), 195–229. https://doi.org/10.1177/1046496403260843

Stroebe, W., & Diehl, M. (1994). Why Groups are less Effective than their Members: On Productivity Losses in Idea-generating Groups. *European Review of Social Psychology*, *5*(1), 271–303. https://doi.org/10.1080/14792779543000084

Suthers, D. D. (2005). Technology Affordances for Intersubjective Learning: A Thematic Agenda for CSCL. In T. D. Koschmann, D. Suthers, T. W. Chan (Eds.), *Computer Supported Collaborative Learning 2005: The Next 10 Years! : Proceedings of the International Conference on Computer Supported Collaborative Learning 2005* (pp. 662–671). Taipei, Taiwan: Lawrence Erlbaum Associates

Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Teasley, S. D. (1995). The role of talk in children's peer collaborations. *Developmental Psychology*, *31*(2), 207–220. https://doi.org/10.1037/0012-1649.31.2.207

Trickett, S. B., & Trafton, J. G. (2007). "What if…": The Use of Conceptual Simulations in Scientific Reasoning. *Cognitive Science: A Multidisciplinary Journal*, *31*(5), 843–875. https://doi.org/10.1080/03640210701530771

van de Pol, J., Volman, M., & Beishuizen, J. (2011). Patterns of contingent teaching in teacher–student interaction. *Learning and Instruction*, *21*(1), 46–57. https://doi.org/10.1016/j.learninstruc.2009.10.004

Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2016). Socio-Cognitive Scaffolding with Computer-Supported Collaboration Scripts: a Meta-Analysis. *Educational Psychology Review*. https://doi.org/10.1007/s10648-016-9361-7

Voss, J. F., Greene, T. R., Post, T. A., & Penner, B. C. (1983). Problem-Solving Skill in the Social Sciences. *Psychology of Learning and Motivation, 17*, 165–213). https://doi.org/10.1016/S0079-7421(08)60099-7

Voss, T., Kunter, M., & Baumert, J. (2011). Assessing teacher candidates' general pedagogical/psychological knowledge: Test construction and validation. *Journal of Educational Psychology*, *103*(4), 952–969. https://doi.org/10.1037/a0025125

Wegge, J., Roth, C., Neubach, B., Schmidt, K.-H., & Kanfer, R. (2008). Age and gender diversity as determinants of performance and health in a public organization: The role of task complexity and group size. *Journal of Applied Psychology*, *93*(6), 1301–1313. https://doi.org/10.1037/a0012680

Weinberger, A., Ertl, B., Fischer, F., & Mandl, H. (2005). Epistemic and social scripts in computer–supported collaborative learning. *Instructional Science*, *33*(1), 1–30. https://doi.org/10.1007/s11251-004-2322-4

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education*, *46*(1), 71–95. https://doi.org/10.1016/j.compedu.2005.04.003

Weinberger, A., Stegmann, K., & Fischer, F. (2010). Learning to argue online: Scripted groups surpass individuals (unscripted groups do not). *Computers in Human Behavior*, *26*(4), 506–515. https://doi.org/10.1016/j.chb.2009.08.007

Weinstock, M. (2009). Relative expertise in an everyday reasoning task: Epistemic understanding, problem representation, and reasoning competence. *Learning and Individual Differences*, *19*(4), 423–434. https://doi.org/10.1016/j.lindif.2009.03.003

Wenglein, S., Bauer, J., Heininger, S., Prenzel, M. (2015). Kompetenz angehender Lehrkräfte zum Argumentieren mit Evidenz: Erhöht ein Training von Heuristiken die Argumentationsqualität? *Unterrichtswissenschaft*, *43*(3), 209–224.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, Modeling, and Metacognition: Making Science Accessible to All Students. *Cognition and Instruction*, *16*(1), 3–118. https://doi.org/10.1207/s1532690xci1601_2

Wiley, J., & Jolly, C. (2003). When two heads are better than one expert. In *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society. Mahwah, NJ: Lawrence Erlbaum Associates Inc*.

Wittwer, J., Nückles, M., & Renkl, A. (2010). Using a Diagnosis-Based Approach to Individualize Instructional Explanations in Computer-Mediated Communication.

Educational Psychology Review, *22*(1), 9–23. https://doi.org/10.1007/s10648-010-9118-7

Wu, D., Liao, Z., & Dai, J. (2015). Knowledge Heterogeneity and Team Knowledge Sharing as Moderated by Internal Social Capital. *Social Behavior and Personality: An International Journal*, *43*(3), 423–436. https://doi.org/10.2224/sbp.2015.43.3.423

Yeh, C., & Santagata, R. (2015). Preservice Teachers' Learning to Generate Evidence-Based Hypotheses About the Impact of Mathematics Teaching on Learning. *Journal of Teacher Education*, *66*(1), 21–34. https://doi.org/10.1177/0022487114549470

Zhao, R., Sinha, T., Black, A. W., & Cassell, J. (2016). Socially-Aware Virtual Agents: Automatically Assessing Dyadic Rapport from Temporal Patterns of Behavior. In D. Traum, W. Swartout, P. Khooshabeh, S. Kopp, S. Scherer, & A. Leuski (Eds.), *Intelligent Virtual Agents* (pp. 218–233). Cham: Springer. https://doi.org/10.1007/978-3-319-47665-0_20

Zimmerman, C. (2000). The Development of Scientific Reasoning Skills. *Developmental Review*, *20*(1), 99–149. https://doi.org/10.1006/drev.1999.0497

Appendix A: Problem Case, Segmentation Scheme, Coding Scheme on Epistemic Activities[2]

I. Problem Case

Stellen Sie sich bitte die folgende Situation vor - wie würden Sie als Lehrer/in handeln, um diese Problemsituation zu beheben?

Sie sind Lehrer/in an einer Schule. Eine Ihrer Schülerinnen erhält oft vergleichsweise schlechte Noten auf ihre Prüfungen. Allerdings erscheint sie im Unterricht sehr motiviert und vermittelt den Eindruck, dass sie fast alle Lerninhalte versteht. Von ihren Eltern wissen sie, dass die Schülerin zu Hause sehr fleißig lernt. Wenn es allerdings zur Prüfung kommt, hat sie allerdings große Schwierigkeiten, das Gelernte abzurufen. Sie als Lehrer/n sollen nun mögliche Gründe für die schwachen Prüfungsleistungen der Schülerin herausfinden.

[2] The segmentation and coding rules have been developed in close collaboration with Christian Ghanem.

II. Segmentation Scheme

a) Grain size

We decided to keep our segmentation relatively fine grain-sized so as to grab the minimum meaningful pieces of information and to follow a micro-level segmentation approach (Weinberger & Fischer, 2006). We chose syntactical unit as the basis of our analysis. Segment borders can be identified by separating propositional units (Chi, 1997, p. 9.) defined by the following rules.

b) Segmentation rules

Segments should be separated by using "//" in the end of each segments.

1. The basic form of a proposition is a full sentence that includes a subject and a statement/verb and is indicated by punctuation signs "." or "?"
2. Complex or compound sentences should be divided into "simple sentences", taking into account the following rules:
   2.1. Everywhere where a comma could be set, a new segment starts, at least if the new sentence part includes a verb.
   2.2. Words like "und", "aber", "oder" typically indicate a new segment, at least if the new and the previous sentence parts contain a verb each.

Example 1:
A: Ok, also wir haben hier die Schülerin, //
die zwar wohl sehr fleißig lernt //
und auch den Eindruck erweckt, //
dass sie alles versteht //
was im Unterricht behandelt wird.//

Exception 1:
"oder so", "klar und offen", "und ding", "weiß ich nicht" (Umgangsprache), "…oder nicht"

2.3. Conditional and other forms of "Nebensatz"-sentences should be handled as compound sentences. Words like "wenn", "ob", "dass", "wann", "wo", "was", "wie", "wer" typically indicate a new segment, <u>at least if the new sentence part contains either a new verb or a new subject or both.</u>

Example 1:
A: Das ist ja eine anerkannte, ähm, Krankheit, //
oder ja, eine eine Angst *(ein unverständliches Wort)* //
oder dass man das halt, ich glaube da man sogar dagegen trainieren oder dass man das halt irgendwie. Ja. //

Exception 1: When "dass" is used without the grammatical value of binding Hauptsatz-Nebensatz.

Example to Exception 1:

A: Aber einfach mal, dass ich mal sehe halt //

Exception 2: Words like "wenn", "ob", "dass", "wann", "wo", "was", "wie", "wer" do not indicate a new segment when they follow an introductory sentence part that expresses a thought, feeling, idea (see rule 4) with phrases like "Er sagte mir, was…" etc.

Exception 3: Simple exemplifying, typically indicated by "z.B." should not be coded as a new segment.

Example to Exception 3:

Genau, man sollte halt eher sich so Eselsbrücken bilden wie jetzt z.B. die Organisationsstrategie? //

3. "ähm", "na ja" within a running sentence should not be coded as a segment on its own.
4. Phrases like "glaube ich", "ich denke", "das heißt", "sie weiß", "es könnte sein", "keine Ahnung", "er sagt", "sie bestätigt", "sie merkt", "das erinnert mich" etc. in a sentence should not be treated as independent segments.
    4.1. Similarly, forms like "das liegt daran", "das hängt davon", "es kommt darauf an", "es wäre gut", "es gibt eine Möglichkeit" as introductory phrases should not be segmented separately.
5. "ja", "genau" or other forms of back channeling, self-reflecting typically should not be coded as different segments. The following rules should apply:
    5.1. Simple backchannels from the learning partner while the other one is speaking should just be ignored (not coded nor segmented).

Example 1:
A: Dass sie dann Verlustängste hat...//
B: Ja.
A: ... wenn es auf den Punkt kommt//

    5.2. When "ja" or "genau" can be considered as backchannels or express agreement and at the same time they are followed by a thought from the same speaker, they should be coded together with the following segment.

Example 1:
Genau. Hatte ich es auch gesagt. //

Example 2:
A: Mhm… genau… Sie lernt ja glaube ich auch sehr viel und fleißig…//

5.3. Similarly, self-reflective use of "genau”, “ja” etc. within a running sentence should not be coded as independent segments.

Example 1:
A: Ja, vor allem es steht ja hier nicht, //
dass sie immer, sondern nur vergleichsweise schlechte Noten, also oft, also, aber, ähm… ja was, genau.//

6. Repetitions of words or sentence parts should belong to the same segment.

Example 1:
A: Das scheint ja auch so… uhm, es scheint ja auch so, //

7. Every new turn except backchannels (Rule 5) should be coded as a new segment regardless how long it is.
8. When no rules from above can apply to separate sentence parts from one another, they should not be divided into separate segments but handled as one.

Additional rules:

9. Sometimes a syntactical structure is interrupted by a parenthetical segment. If the parenthetical qualifies as a full segment, we should segment it separately from the interrupted syntactical structure.

   Moreover, if the interrupted structure qualifies on its own (without the parenthetical) as a full segment, we should segment both the part before and after the parenthetical structure, separately.

   Still, in an additional comment "[s+-x]" we should make it clear that the separated pieces of the interrupted segment belong together, and they should be handled as one segment. In the comment "[s+-x]" x represents the number of steps (+ or -) where the corresponding segment part occurs. Consequently, [s+2] means that the corresponding segment part is a subsequent one, in a two-step distance from the actual segment part, while [s-2] means, the corresponding segment part precedes with 2 steps the current segment part.

   Example 1:
   B.: Und da haben halt, // [s+2]
   je nachdem wie alt das Kind ist halt, //
   die Eltern auch einen großen Einfluss drauf.//[s-2]

III. Coding scheme

a) Relationship between segments and codes

Since our choice was a relatively small segment size, it is possible that one code can be assigned throughout multiple segments. In this case all the segments can be considered under one code if they are a continuation of the first segment where the same code was assigned (see Example 1 for instance).

b) Epistemic activities of scientific reasoning

The epistemic activities of scientific reasoning are suggested by Fischer et al. (2014) as they intend to comprehensively enough cover those actions a reasoner might take during scientific reasoning. Our coding scheme targets the question how and to what extent these activities can describe teacher students' scientific reasoning during solving a practice-related task.

1. Problem identification (PI)

Criteria:
PI should be coded in case of a reference to the problem description that expresses

a) An attempt to build an understanding of the problem described
b) An actual understanding of the problem (like summarizing the problem description)

and it should precede further inquiry. Therefore, PI has a sequential characteristic, typically occurring in the beginning of scientific reasoning.

Clarification:
This epistemic activity is related to "Construction of problem space" in the Weinberger & Fischer (2006) coding scheme: during PI participants deal with the available case information in the task description with the epistemic aim to build a problem representation. They do so, so as to identify how to direct their further inquiry. Participants notice that the problem cannot be answered without further inquiry. To reach the epistemic aim of understanding the problem better, participants might:

1. read the case information (task description) out loud
2. connect case information with case information within the task description
3. reflect on the case information (rephrase, interpret, ask questions on the case information)

It is also possible, that participants make generalizations on the case information (similarly to Chi et al, 1981), but this should be part of their understanding process so as it is not coded otherwise.

Examples:

Example 1.

„Ok, also wir haben hier die Schülerin, //
die zwar wohl sehr fleißig lernt //
und auch den Eindruck erweckt //
dass sie alles versteht //
was im Unterricht behandelt wird. //
Allerdings, uhm, kann sie dann in den Prüfungen das Gelernte nicht abrufen //"

Rule 1: Reference to the task description in the very beginning of the transcript is typically coded under PI while later references are always considered under EG or EE as case evidence. So PI should almost always be coded in the beginning but never afterwards.

2. Questioning (Q)

Criteria:
This epistemic activity is orienting further inquiry by focusing the interest on a main question that should be answered in order to solve the problem. It is notable that Q does not have to have a question format (see Example 3).

Clarification:
It is an initial question coming from the original problem (described in the problem scenario and might be explicitly referred during PI) and it directs subsequent reasoning processes. It comes from the situation where the problem exists and it targets further information collection / analysis / discussion etc.

This epistemic activity might - although not necessarily - follow problem identification. "Based on the representation developed during problem identification, one or more initial questions are identified for the subsequent reasoning process." (Fischer et al, 2014., p...)

Examples:

Example 2.
"O.k., warum ist sie so schlecht in den Prüfungen? //"

Example 3.
A: „Ähm, ich glaub wir sollen da jetzt eine Lösung dazu finden.//
B: Dass sich praktisch die Prüfungsleistungen verbessern, //
oder.. ne, mögliche Gründe eher.//
A: Also mögliche Gründe...
B: mhm
A: ...wieso sie quasi die, ähm, Leistung die sie im Unterricht eben nicht abrufen kann. //"

Rule 1: Just like PI, Q is typically coded in the very beginning in the transcript. In contrast to PI, it is possible to code Q later but only if it leads to further inquiry.

3. Hypothesis generation (HG)

Criteria:
It is a plausible (and ideally testable) explanation of the problem, and it should be a plausible answer to Q, although it does not matter if PI or Q is explicitly stated by the participant or not.

Clarification:
During this epistemic activity participants normally “derive possible answers to the question from plausible models, available theoretical frameworks or empirical evidence they are aware of” (Fischer et al. 2014, p...). It can come from theories that can give plausible answers: for instance, "Lernstrategie" as explaining "Prüfungsproblem". However, it does not have to be scientific. Therefore, any possible explanation of the problem can qualify as HG.

Examples:

Example 4.
"A: […] Oder, wie du schon gesagt hast, Prüfungsangst…//
[reden durcheinander]
B.: Ja, also entweder es ist halt Prüfungsangst, //
dass sie es halt eigentlich kann //
aber dann
A.: Genau
B.: so gestresst ist
A.: Ja
B.: in der Prüfung, //"

Example 5.
"Oder vielleicht lernt sie’s….vielleicht hat sie auch die falsche Lernstrategie. //"

Example 6.
"Ja, vielleicht kann sie auch einfach nicht das, was sie lernt, //
mit anderen Sachen verknüpfen. //"

Rule 1: Hypothesis should not be confused with hypothetical EG or hypothetical EE. A hypothesis is always a possible reason of the student’s problem. A hypothetical evidence is typically a “what if” evidence that could happen in a hypothetical scenario but it does not serve as a possible reason for the student’s problem.

4. Construction and redesign of artefacts (CA)

Criteria:

We consider artefact as an intervention / solution plan. This intervention plan can be formed as a solution to the original problem or they my aim changing the context of the problem to reach a solution.

Clarification:

The artefact is typically a plan for revision of the context where the problem occurred so as to prevent the problem or to solve it. It can be:

- teaching methods
- curriculum design
- testing methods
- other intervention plan where an artefact (e.g. training session) can be identified (see Example 6.)

CA differs from HG, because it does not only say a plausible solution but the epistemic aim is to restructure the problem context itself in a way that the problem is alleviated, solved, prevented.

Examples:

Example 6.
"Vielleicht kann man auch noch irgendwie externe Ressourcen einbringen, wie z.B. ..ja, Nachhilfe, //"

5. Evidence generation (EG)

Criteria:

This epistemic activity should be coded when participants use any available information or generate new (e.g. hypothetical) information that could help in solving the problem or answering the question.

Clarification:

We consider evidence as any form of information taken into account with the aim to support or to reject a claim or to make more reliable decisions.

Typical types of evidence:

1. **Case information** written in the problem statement
2. **Scientific content** information on the Folien
3. Scientific content information other
4. Nonscientific, **anecdotal** information (personal experience) retrieved from memory
5. **Hypothetical** information ("what if")

We consider hypothetical evidence as a valid form of "scientific" evidence in the process of "what if" reasoning or "conceptual simulation": as Trickett & Trafton (2007) claims (p. 844) that in case of informational uncertainty when for instance data is missing, scientists can generate it by thinking of alternatives. The authors consider it as a valid form of reasoning process. We understand hypothetical evidence in this sense.

Examples:

Example 7.
"Aber die Sache ist ja hier, hier steht ja, //
dass sie fast alle Lerninhalte versteht. //" (EG type 1.)

Example 8.
"Aber dann würde sie sich am Unterricht nicht so beteiligen //
und dann würde der Lehrer nicht sagen sie kann's. //" (EG type 5.)

Example 9.
"ob's in den anderen Fächern auch so ist.//" (EG type 5.)

Rule 1: When evidence is used with evaluative purpose, it should always be coded under EE with a reference to what type of evidence it is (scientific, case, anecdotal, hypothetical).

Rule 2: When a segment cannot be coded under any types of the evidences above, it is not considered evidence and should not be coded as EG or EE.

6. Evidence evaluation (EE)

Criteria:

Participants deal with the evaluation of available or generated information; or with the hypothesis.

Clarification:

This epistemic activity typically represents the phase of inquiry that deals with the relationship between a claim (hypothesis, theory, model, initial question) and any available or generated information. According to Sandoval et al (2014) "both the claim itself must be evaluated, and evidence is evaluated in relation to the claim." (p 141.). However, our interpretation is wider than that: we suggest that any evaluation of a claim or evidence regarding to its reliability or validity can also be considered under EE.

1. Participants evaluate evidence if it supports or falsify claims or any pre-existing ideas (theories, hypotheses, initial questions).
2. Participants evaluate a claim whether it can be kept or needs to be revised. Typically hypothesis falsification (even without a reference to the evidence). (See Example 10.)
3. Participants evaluate the reliability of the evidence or its source.
4. Participants deal with the validity (typically contextual nature) of the problem, evidence or a hypothesis (Example 11).
5. Evaluative statements on problem, evidence or hypothesis are also EE.

Examples:

Example 10.
„Steht ja, glaube ich, da irgendwo, //
dass das nicht da drauf ankommt, dass man es lernt //“

Example 11.
„Ja, das ist ja bei einer bestimmten Situation so“

Rule 1: EE should only be coded in case of one of the 5 types described above. If the segment cannot be coded under these, that is either another activity or non-epistemic.

7. Drawing conclusion (DC)

Criteria:

Participants make conclusive statements regarding the inquiry process, e.g., to summarize.

Clarification:

What follows from the results? Can we say anything, plan intervention etc? This activity is more conclusive than EE (does not deal only with the quality of evidence) or HG (it is not "only" an answer to Q). It deals with the outcomes of the reasoning process, the summary and further plans maybe.

Examples:

Rule 1: DC is always summative. It typically has sequential characteristic to happen in the end of the reasoning cycle where participants summarize their findings.

Rule 2: It can happen not at the end of the transcript, but earlier, in case participants finish thinking about the reasons of the problem and start with the solutions. Still, it has to be at least a brief summary of the previous ideas.

8. Communicating and scrutinizing (CS)

Criteria:

This epistemic activity is coded when participants plan to engage in communication with the aim of developing scientific inquiry, regardless of what stage of it.

Clarification:
Communicating might happen in order to generate evidence, evaluate evidence or hypothesis, construct artefacts, draw conclusions and so on. The main criteria is the plan to involve others in the inquiry process with the epistemic aim to develop a better (more reliable) outcome.

Examples:

“Vielleicht könnte man da mit der Schülerin dann einfach drüber reden //“

“Vielleicht sollen wir auch mit anderen Lehrern diskutieren. // “

Rule 1: CS should be coded only if there is a plan to communicate and to develop scientific reasoning at the same time. If the plan is only to talk without the aim to collect evidence, develop intervention and so on, then it should not be coded as CS.